The diverse roles of the \textit{Pseudomonas aeruginosa} phospholipase PlcH and the \textit{Toxoplasma gondii} serine palmitoyltransferase in sphingolipid biochemistry.

THYE, JULIE,KRISTINE

The diverse roles of the *Pseudomonas aeruginosa* phospholipase PlcH and the *Toxoplasma gondii* serine palmitoyltransferase in sphingolipid biochemistry

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in the University of Durham by

Julie Kristine Thye

July 2014

Department of Chemistry

University of Durham, Van Mildert College

Abstract

Sphingolipids are essential for cell survival for almost all eukaryotic cells. Due to the large variety in sphingolipid structure, they possess a wide range of functions from acting as structural components of membranes to participating in cell signalling pathways. Sphingolipid metabolism and catabolism can occur via numerous routes and the disruption of sphingolipids or sphingolipid biosynthetic pathways is therefore detrimental to cells.

The vast majority of prokaryotes do not utilise sphingolipids, however certain prokaryotic enzymes which posses similar functionality to those involved in sphingolipid biosynthesis, are crucial to pathogen-host interactions. For instance, the opportunistic human pathogen, *Pseudomonas aeruginosa*, which represents a particular threat to cystic fibrosis patients, expresses a variety of virulence factors. One of these, the heterodimeric complex, $PlcHR_2$, is a phospholipase C. It also possesses sphingomyelinase properties which has been suggested to be responsible for the highly specific cytotoxic effects of the enzyme. The first aim of the thesis was to express and purify the PlcH part of the complex in order to functionally and biophysically characterise it, laying the foundation towards the crystal structure determination. Furthermore, the sphingomyelinase activity assay was established, utilising $PlcHR_2$, in order to screen several commercially available compounds and an in-house library of ceramide analogues for potential inhibitory effect on the enzyme. Five inhibitory and five activating compounds were identified.

Sphingolipids also play crucial roles in more complex human pathogens, including the apicomplexan protozoan parasite, *Toxoplasma gondii*, which causes a life-threatening disease, toxoplasmosis, in animals and humans. The first step in the sphingolipid biosynthetic pathway is catalysed by serine palmitoyltransferase (SPT), making it an important drug target in several parasites as preventing this step from occurring in the biosynthesis will result in an incapability of the parasite to proliferate. Several constructs of the SPT from *T. gondii* were designed in order to establish the smallest catalytically active domain. The constructs were expressed and purified in order to characterise the enzyme by a range of biochemical and biophysical methods, with the ultimate aim of obtaining a crystal structure for the enzyme.

The final part of the thesis focussed on the crystallisation and crystal structure determination of a triosephosphate isomerase (TPI) from a hyperthermophilic archaeon, *Sulfolobus solfataricus*. The TPI protein assumes a $(\beta\ \alpha)_8$-barrel fold, the most common protein fold. It is thus an ideal target to study to gain an understanding of the evolutionary process that enzymes undergo to stabilise their structure and function at the extreme conditions they are subjected to in extremophiles.

Acknowledgements

First and foremost I would like to thank my supervisor, Dr Ehmke Pohl, for all his guidance, support and patience throughout my project and for providing me with many opportunities to learn a range of techniques. My co-supervisor, Dr Paul Denny, has been very helpful regarding the biology aspects of this thesis.
I would also like to thank Diamond Light Source for providing half my funding and Dr Martin Walsh and the entire MX team for providing excellent beam line facilities.
I am very grateful to the Dr Ray Owen, Dr Louise Bird, Dr Joanne Nettleship and Dr Anil Verma at the Oxford Protein Production Facility for providing all the help and suggestions needed with cloning, protein expression and purification as well as protein sample mass spectrometry and crystallisation tray assistance.
The protein samples provided by Dr Michael Vasil (University of Colorado, Denver), have been invaluable and without them would never have gotten a decent amount of data for my PlcH chapter.
Many thanks to Dr Dominik Esser and Dr Bettina Siebers (Department of Chemistry, Molecular Enzyme Technology and Biochemistry, at the University of Duisburg-Essen, Germany) for sending the samples of pure *Ss*TPI, providing me with the essential part for a protein crystallography project - crystallisable protein.
Ian Edwards has been an invaluable help, teaching me to use the FPLC, helping with crystallisation trays and finding my way round the lab. I would like to thank Dr Morten Grøftehauge for his ideas and help with the thermostability assay, X-ray data collections as well as the SAXS data collection and analysis. Thanks to Dr John Mina who has been invaluable for discussion regarding the sphingomyelinase assay. A special thanks to Robert Dods for ensuring that the SPT project got off to a running start and for providing me with entertainment throughout the second year blues, and Laura Dowd for being such a quick learner and preparing extra samples for the SAXS experiments.
I would also like to thank all members of office CG231 and Lab229, past and present, for all the enjoyable tea-breaks and random lab chats. I am also grateful to Alba, Aruna, Frances, Jenny and Zara for being there when things did not do what they were meant to.
Away from the laboratory I would like to thank all my friends from home and undergraduate studies. Especially Helen, who has been there since the beginning and has provided great company during many lunch breaks. Thank you Hayley for never turning down a catch-up and being a great friend. Thank you to the Luxembourg crowd, especially Anne-Marie, Anoushé and Laura, for funny coffee meet-ups and catch-ups whenever we all manage to be in the same country at the same time, and Emily who has suffered with me these last few weeks but still kept the motivation

going.
Thank you to John for listening to all my stress, supporting me during the most difficult days without judgement, and for always being there. Last but not least I am so grateful to my family, my parents and my sister for all their advice, help and never-ending cheering and motivation from home.

Declaration

The work presented herein was carried out in the Department of Chemistry at Durham University between October 2010 and March 2014. In January and February 2011, work was conducted on PlcH at the Oxford Protein Production Facility over two weeks. Unless otherwise stated, all work is my own and has not been submitted previously for a qualification at this or any other university.

Statement of Copyright

Contents

List of Figures

List of Tables

Abbreviations

AIDS	acquired immunodeficiency syndrome
AOS	α-oxoamine synthase
asu	asymmetric unit
ATP	adenosine triphosphate
BBB	blood-brain barrier
Bis-Tris	bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane
bp	base pairs
C1P	ceramide-1-phosphate
CerS	ceramide synthase
CF	cystic fibrosis
CFTR	cystic fibrosis transmembrane conductance regulator
CMC	critical micelle concentration
CNS	central nervous system
Da	Dalton
DAT	direct agglutination test
DCS	D-cycloserine
DHAP	dihydroxyacetone phosphate
DLS	Diamond Light Source
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNAse	deoxyribonuclease
DSF	differential scanning fluorimetry
DTT	dithiothreitol
ED	Entner-Doudoroff
EDTA	ethylene diamine tetraacetic acid
EM	electron microscopy
EMP	Embden-Meyerhof-Parnas
ER	endoplasmic reticulum
ESI	electrospray ionisation
FA	fatty acyl
FPLC	fast protein liquid chromatography
GAP	D-glyceraldehyde-3-phosphate
GL	glycerolipid
GlcCer	glucosylceramide
GlcCerS	glucosylceramide synthase

GP	glycerophospholipid
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
HRV3C	Human rhinovirus 3C
HTP	high-throughput
IFA	indirect fluorescent antibody
IMAC	Immobilised Metal Affinity Chromatography
IPC	inositol phosphorylceramide
IPTG	isopropyl β-D-1-thiogalactopyranoside
KDS	3-ketodihydrosphingosine
LB	lysogeny broth
LCS	L-cycloserine
LPS	lipopolysaccharide
MALDI-ToF	matrix assisted laser desorption/ionisation time-of-flight
MeOH	methanol
MES	2-(N-Morpholino)ethanesulfonic acid
mins	minutes
MOPS	3-(N-Morpholino)propanesulfonic acid
MPD	2-methyl-2,4-pentanediol
MR	molecular replacement
MV	membrane vesicle
MW	molecular weight
MWCO	molecular weight cut-off
NE	neutrophil elastase
Ni-NTA	nickel-nitrilotriacetic acid
NMR	nuclear magnetic resonance
OCD	obsessive compulsive disorder
OPPF	Oxford Protein Production Facility
PAGE	polyacrylamide gel electrophoresis
PC	phosphatidylcholine
PC-PLC	phosphatidylcholine phoshpolipase C
PCR	polymerase chain reaction
PEG	polyethylene glycol
PEP	phosphenol pyruvate
Pi	phosphate
PI	phosphatidylinositol
PIPES	1,4-Piperazinediethanesulfonic acid
PK	polyketide
PKA	protein kinase A
PLC	phospholipase C

PLP	pyridoxal 5'-phosphate
PMP	polymethylpentene
POAS	PLP-dependent α-oxoamine synthase
PPMP	1-phenyl-2-palmitoylamino-3-morpholino-1-propanol
PR	prenol lipid
PS	phosphatidyl serine
PV	parasitophorous vacuole
RNA	ribonucleic acid
RNAse	ribonuclease
S1P	sphingosine-1-phosphate
SAR	structure-activity relationship
SAXS	small-angle X-ray scattering
SDS	sodium dodecylsulphate
secs	seconds
SL	saccharolipid
SM	sphingomyelin
SMase	sphingomyelinase
SMS	sphingomyelin synthase
SP	sphingolipid
SPT	serine palmitoyltransferase
ST	sterol lipid
SUMO	Small Ubiquitin-like Modifier
SUV	small unilamellar vesicle
Tat	twin arginine translocase
TBonex	Overnight Express^{TM} Instant TB Medium
TCA	trichloroacetic acid
TCEP	tris(2-carboxyethyl)phosphine
TDTMA	tetradecyltrimethylammonium
TFA	thermofluor assay
TLC	thin-layer chromatography
TPI	triosephosphate isomerase
Tris	tris(hydroxymethyl)aminomethane
TSA	thermal stability assay
UDP-galactose	uridine diphosphate galactose
WT	wildtype
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Chapter 1

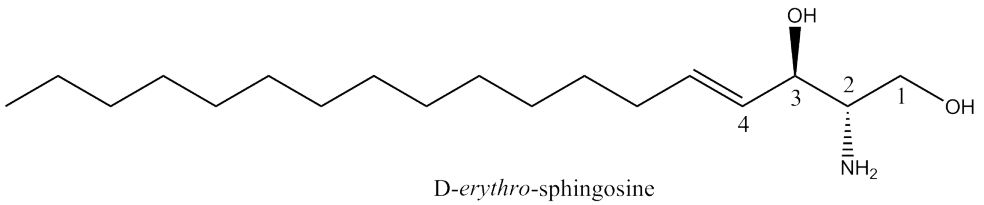
The biochemistry of sphingolipids

1.1 Introduction

Sphingolipids are a specific subclass of lipids which play many important roles within cells being both involved in the physiological as well as the pathophysiological aspects.[1,2] Sphingolipids are important structural components of eukaryotic cell membranes but are also involved in a number of intracellular processes.
In order to survive and proliferate, parasites rely on complex lipid metabolic pathways to aid rapid cellular division. These pathways are catalysed by a vast variety of enzymes, many of which have been identified as potential drug targets.[3,4]

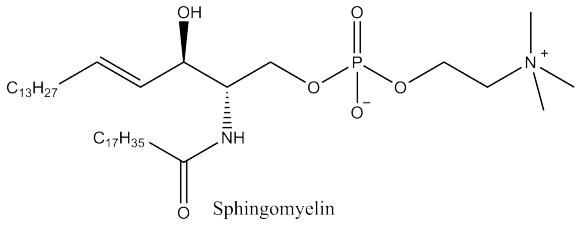
1.1.1 Sphingolipid structure and classification

Sphingolipids consist of a long-chain base (also known as a sphingoid base) backbone such as sphingosine ((2*S*,3*R*,4*E*)-2-aminooctadec-4-ene-1,3-diol)[5] (Figure 1.1), the most common sphingoid base in mammals.[6,7] To this backbone, several moieties - neutral, charged, phosphorylated and/or glycosylated - can be attached to form the complex sphingolipid.[7] Due to the difference in polarity in the molecules arising from the various moieties and long-chain fatty acid, most sphingolipids are amphipathic in nature.
Sphingolipids were first discovered by J. L. W. Thudichum in 1884,[8] who isolated three compounds; sphingomyelin (SM), cerebroside and cerebrosulphatide (Figure 1.2) from human brain tissue. They belong to the three known types of sphingolipids[9] namely the phosphosphingolipids (Figure 1.2 a), neutral glycosphingolipids (Figure 1.2 b) and acidic glycosphingolipids (Figure 1.2 c).[10]

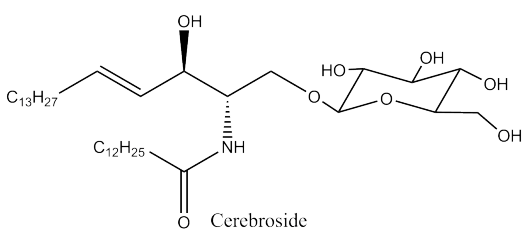


D-*erythro*-sphingosine

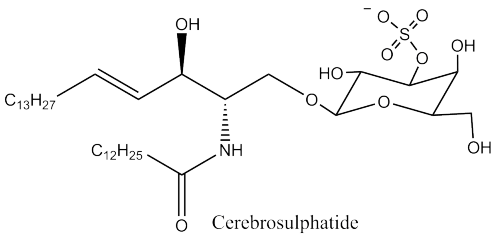
Figure 1.1: The chemical structure of sphingosine, also known as D-*erythro*-sphingosine, (2*S*,3*R*,4*E*)-2-aminooctadec-4-ene-1,3-diol or sphing-4-enine.



(a)



(b)



(c)

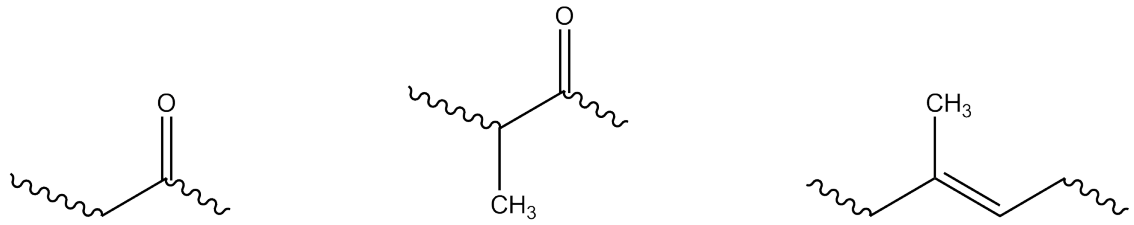
Figure 1.2: The chemical structures of a) sphingomyelin; b) cerebroside and c) cerebrosulphatide.

Due to the large number of processes and biological functions that sphingolipids are involved in (section 1.1.3), this area of biology has been growing rapidly in the last decade.[6,11]

1.1.1.1 Lipid structure and classification

Lipids have been defined as small hydrophobic molecules insoluble in water but soluble in organic solvents.[10,12] The huge variety in structure and functionality of lipids means that a comprehensive classification system is necessary in order to define each lipid. Current nomenclature and classification of the various lipids is based on the classification outlined by the Lipid Metabolites and Pathways Strategy (LIPID MAPS) Consortium.[13] More than 37,000 lipids are classified out of which

more than 4000 are sphingolipids. The common notation is based on the notion that lipids are constructed from ketoacyl and isoprene as building blocks assuming that all lipids are products of condensation reactions of ketoacyl thioesters and/or isoprene unit condensations.[12] Chain extensions with ketoacyl units are based on a Claisen condensation which occurs via a carbanion intermediate and chain extension with isoprene units is an alkylation reaction which occurs via a carbocation intermediate. The building block structures are shown in Figure 1.3.



(a) (b) (c)

Figure 1.3: The chemical structures of two ketoacyl building blocks a) acetyl; b) propionyl and c) an isoprene building block.[12]

The classification system is based on the subdivision into eight categories according to their structure. These categories are the fatty acyls (FA), glycerolipids (GL), glycerophospholipids (GP), sphingolipids (SP), saccharolipids (SL), polyketides (PK), sterol lipids (ST) and prenol lipids (PR), as shown below in Figure 1.4.[12]

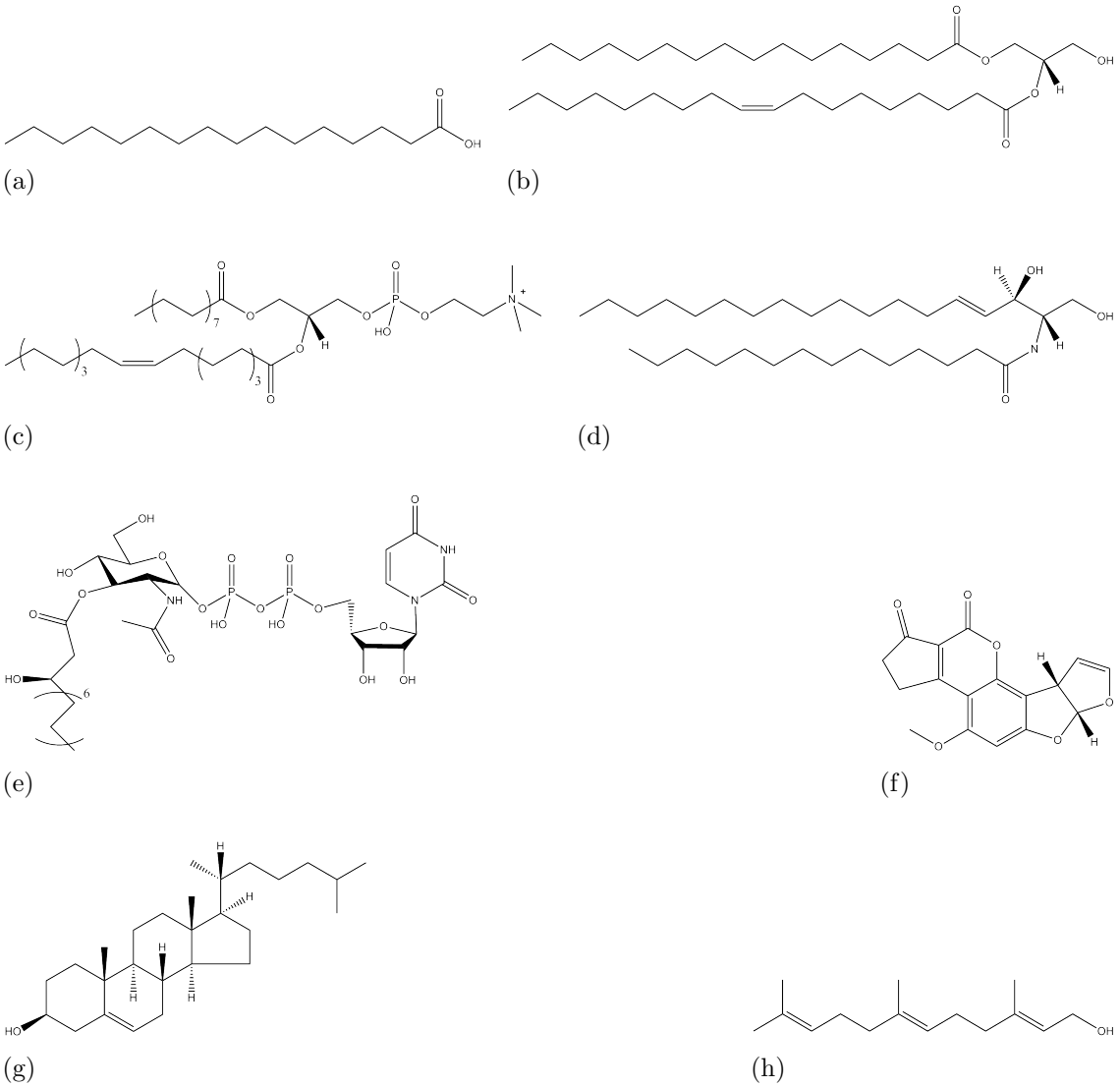


Figure 1.4: Structures of representative examples of the eight LIPID MAPS lipid categories. a) Fatty Acyls: hexanedecanoic acid; b) Glycerolipids: 1-hexadecanoyl-2-(9*Z*-octadecenoyl)-*sn*-glycerol; c) Glycerophospholipids: 1-hexadecanoyl-2-(9*Z*-octadecenoyl)-*sn*-glycero-3-phosphocholine; d) Sphingolipids: *N*-(tetradecanoyl)-sphing-4-enine; e) Saccharolipids: UPD-3-*O*-(3*R*-hydroxy-tetradecanoyl)-αD-*N*-acetylglucosamine; f) Polyketides: aflatoxin B_1; g) Sterol Lipids: cholest-5-en-3β-ol and h) Prenol Lipids: 2*E*,6*E*-farnesol. Each of these classes can be further divided into subclasses.

From the eight categories, the sterol and prenol lipids are derived from isoprene building blocks whereas all the other six categories are biosynthetically based on condensation reactions between ketoacyl building blocks. The glycerolipids, glycerophospholipids and the sphingolipids all contain long chain fatty acid backbones with various moieties attached, leading to their unique functionalities.

1.1.1.2 Sphingolipid structure

The sphingolipid family has a common structure, namely a sphingoid base backbone to which amide bonds link various moieties.[14] Sphingoid bases are structurally very similar to sphingosine.[15] They represent a large variety of 2-amino-1,3-dihydroxyalkanes and -enes; alkyl chains varying between 14 to 22 carbons; 0, 1 or 2 double bonds and (2*S*,3*R*)-erythro stereochemistry. At the 4 and 6 position, hydroxyl groups and other modifications are generally found and at either ω-1 (iso), ω-2 (anti-iso) or at another position, branching methyl groups are often located.[16] The most common sphingoid bases are sphinganine or sphingosine;[2,15] the difference between the two consists of a *trans* 4 - 5 double bond in sphingosine, which affects processes the lipid is involved in.
The sphingolipid family is further subdivided into multiple major classes. These are the sphingoid bases and the simple derivatives thereof such as 1-phosphate; sphingoid bases which contain an amide-linked fatty acid such as ceramides; and the complex sphingolipids which contain phosphodiester-linked head groups such as phosphosphingolipids; head groups attached by glycosidic bonds (the glycosphingolipids which can be subdivided into the simple (cerebrosides) and the complex (gangliosides) glycerosphingolipids); and last but not least the arseno- and phosphonosphingolipids which contain other head groups.[14]

1.1.2 Metabolic pathways of sphingolipids

The biosynthesis of sphingolipids is highly complex and involves a number of intersecting and branching pathways.[6] Synthesis and degradation typically occurs in several intracellular sites such as the endoplasmic reticulum (ER), the Golgi apparatus or in lysosomes.[17,18] The sphingolipid metabolic pathway has been of special interest to scientists for the last few decades resulting in the characterisation of most enzymes involved in the specific steps.[2] Due to the nature of the lipid products of the pathway, most of the enzymes catalysing the various reactions are integral membrane proteins.
Ceramide is the central compound in the sphingolipid metabolism as it links both synthesis and catabolism.[1] In Figure 1.5, a small part of the eukaryotic sphingolipid metabolism is summarised, showing the first steps towards ceramide synthesis and the branch points from this central position.

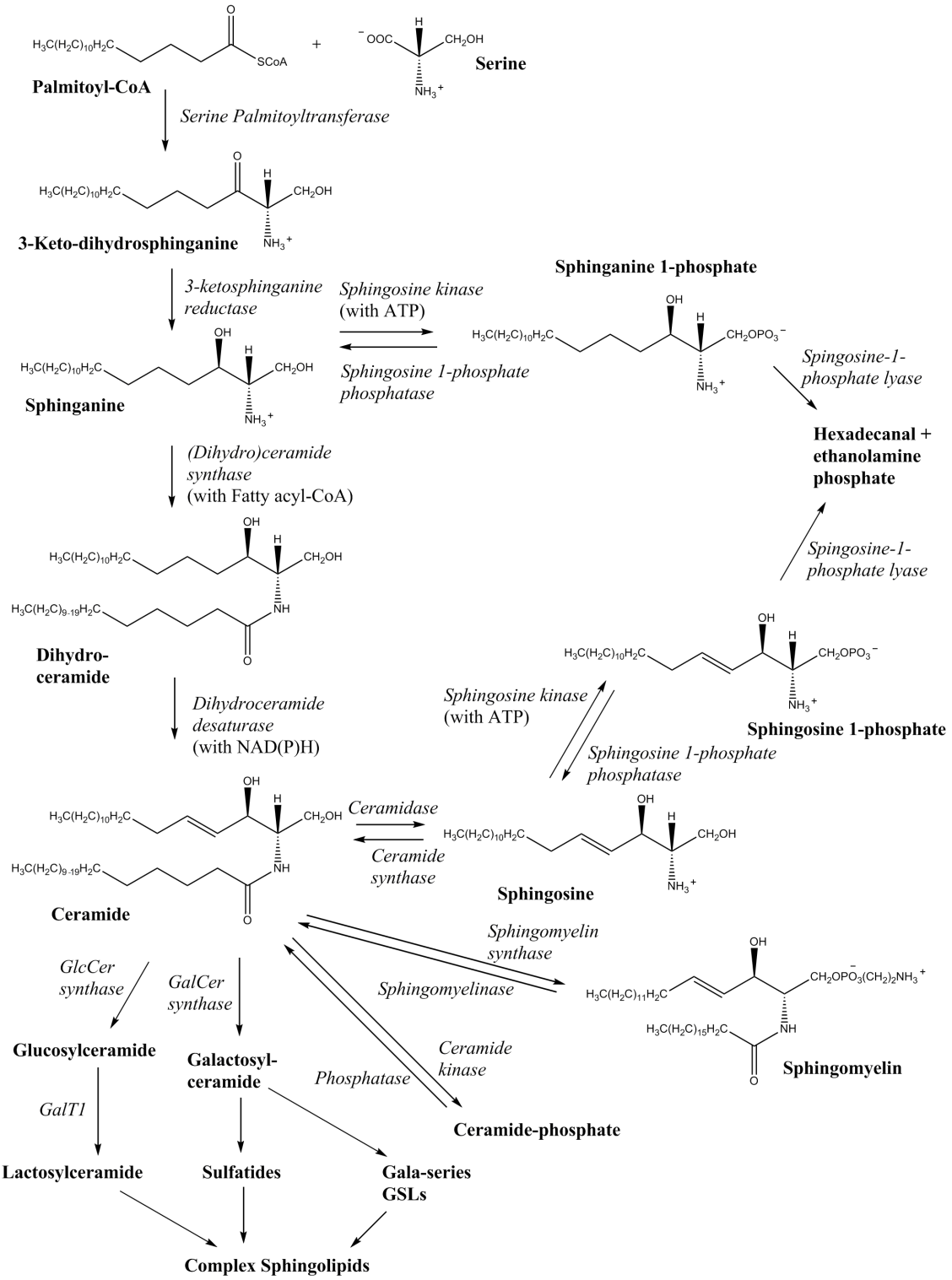


Figure 1.5: The sphingolipid metabolic pathways in mammalian cells. Words in bold are product compound names, words in italics are the names of enzymes catalysing specific reactions. Adapted from Figures 1 and 2 by Merrill[15] and Lahiri *et al*,[2] respectively.

1.1.2.1 Ceramide - a central compound in the sphingolipid metabolism

Ceramide, the simplest sphingolipid, is the central molecule in the sphingolipid metabolic pathway[2] as it acts as a precursor for the complex sphingolipids as well as playing an important role as a signalling molecule, regulating cellular functions.[1,15,19,20,21] Ceramides are a subclass of the sphingoid bases, which has an amide-linked fatty acid and are thus known as N-acyl-sphingoid bases. The fatty acid can be both mono-unsaturated or saturated and the chain lengths are generally between 14 and 26 carbons. A wide range of processes involve ceramide, such as cell growth, apoptosis, neurodegenerative disorders, inflammation, insulin resistance and diabetes to name a few.[1,21,22] It is typically associated with apoptosis and blocking the cell cycle.
Ceramide can be generated by *de novo* synthesis, by recycling other lipid components within the cell or by degradation of more complex sphingolipids,[23] but it has still not been determined how it is split between separate pools for the various functionalities even though the separation into specific pools for cell usage is vital.[21]

1.1.2.2 *De novo* sphingolipid biosynthesis

De novo synthesis of sphingolipids has been discovered in eukaryotes including fungi as well as in bacterial species. The first two steps in the metabolism of sphingolipids is highly conserved across the eukaryotes capable of *de novo* sphingolipid biosynthesis.[2,24] The first step is the condensation reaction of palmitoyl-CoA and L-serine to form 3-ketodihydrosphingosine (KDS, also commonly referred to as 3-ketosphinganine), catalysed by serine palmitoyltransferase (SPT).[25,26,27] The second step consists of the reduction of KDS by 3-ketosphinganine reductase, resulting in D-*erythro*-sphinganine. These two steps are shown below in Figure 1.6.
The first step is the rate-limiting step to all sphingolipid synthesis, making SPT a remarkable drug target as loss of this mechanism without an outside sphingolipid source will prevent normal cellular function.[15]

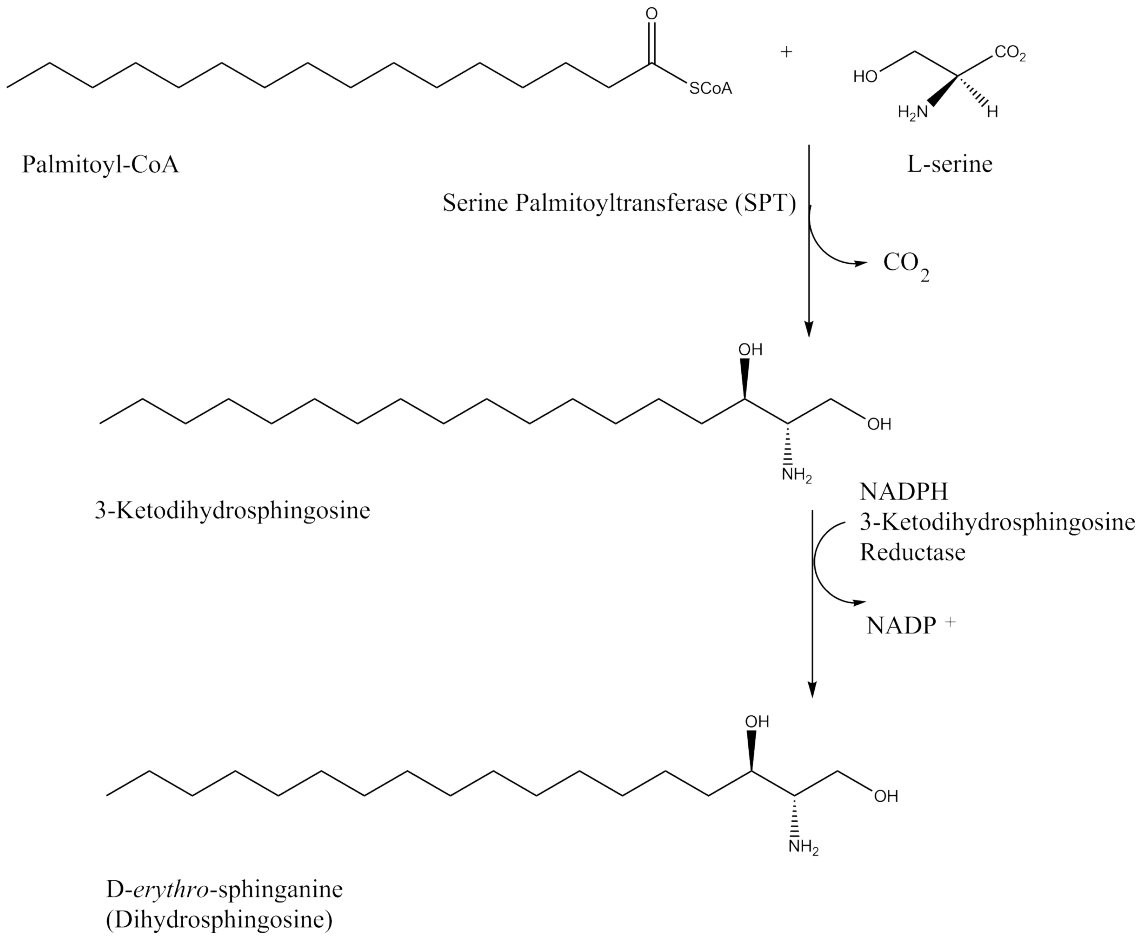


Figure 1.6: Diagram of the synthesis of D-*erythro*-sphinganine in mammalian systems.

Ceramide defines a central point in the sphingolipid metabolic pathway as it functions as a precursor for at least five different products.[2] Sphinganine is a precursor to ceramide and it can be acylated by dihydroceramide synthase (which is known as CerS or Lass) resulting in the formation of dihydroceramide (Figure 1.5). Following the pathway outlined in Figure 1.5, dihydroceramide desaturases/reductases result in a *trans* 4 - 5 double bond insertion in dihydroceramide to form ceramide at the cytosolic side of the ER. (i) Galactosylceramide (GalCer) can be synthesised by glycosylation of ceramide by a UDP-galactose donor transferring galactose to ceramide, which takes place at the ER's lumenal leaflet, catalysed by galactoceramide synthase (GalCerS). (ii) Ceramide-1-phosphate is formed when a ceramide kinase phosphorylates ceramide. (iii) Deacylation by ceramidases of ceramide can result in sphingosine and free fatty acids.

The various members of the CerS family have different preferences for the subsets of acyl CoAs utilised for synthesising dihydroceramide resulting in specific ceramides, which in turn possess different properties as this is linked to regulation of differ-

ent cell functions[23] and therefore localise in different parts of the cell.[2] They are thus distributed in the tissue to optimise synthesis. CerSs are divided according to which pH they function at. Neutral CerSs are found in the plasma membrane (PM), acidic ones are associated with lysosomes[2] and alkaline CerSs are generally found at the ER/Golgi complex. (iv) Sphingomyelin can be formed in the Golgi apparatus from ceramide via an SM synthase when phosphocholine is transferred from phosphatidylcholine to ceramide by sphingomyelin synthase (SMS) resulting in ceramide and diacylglycerol (DAG).[1,4] (v) Last but not least, ceramide can be glycosylated by glucosylceramide synthase (GlcCerS) in the Golgi apparatus, resulting in glucosylceramide (GlcCer).[4]
One process which influences the ceramide levels is the degradation of it to sphingosine by particular ceramidases whereafter sphingosine can be phosphorylated to form sphingosine-1-phosphate (S1P).
As sphingomyelin synthase (SMS) regulates both ceramide and diacylglycerol (DAG) levels in cells, it has been linked to both cell regulation and transformation, and it is also believed to affect processes which have usually been linked to phosphatidylcholine specific phospholipase C.[28]

1.1.2.3 Catabolism of sphingomyelin to ceramide

The intracellular levels of sphingolipids are also maintained by recycling and degradation pathways where ceramide is regenerated from more complex sphingolpids such as sphingomyelin (SM), sphingosine or glycosphingolipids.[1] This can occur via reservoirs of complex sphingolipids that are catabolised by hydrolases, forming glycosylceramide or galactosylceramide whereafter β-glycosidases and galactosidases cleave off the sugars resulting in ceramide. This is known as lysosomal degradation. An alternative is an SMase catabolic pathway, known as non-lysosomal degradation.[23,29] Sphingomyelinase (SMases) (neutral, acidic or alkaline) are phosphodiesterases and they catalyse the catabolism of SM to ceramide and phosphocholine, whereafter the produced ceramide can be further degraded to sphingosine or can be utilised for further sphingolipid formation. This process can lead to apoptosis which has been observed during infection with the opportunistic pathogen, *P. aeruginosa*. It has been associated with the enzyme $PlcHR_2$ which possesses SMase activity and which has been linked to the mortality from *P. aeruginosa* infections due to poor wound healing and vasculitis, as $PlcHR_2$ is lethal to endothelial cells.[30]
Cellular ceramide levels are also affected by the sphingosine salvage pathway where sphingosine can be recycled to ceramide as a result of ceramide synthases.[1]

1.1.3 Biological function of sphingolipids

Sphingolipids have numerous functions both structurally as well as involvement in intracellular processes, as outlined briefly below.

1.1.3.1 Structural components of membranes

Eukaryotic cell membranes consist of a complex lipid bilayer usually containing lipids from three main classes; glycerolipids, sterols and sphingolipids, which form the outer leaflet of the plasma membrane lipid bilayer.[2,4] These lipids provide structure and architectural stability to the cell due to their bonding abilities associated with the difference in polarity across the molecules. Moreover, they play a role in chemical resistance as the lipid bilayer protects against harmful substances.

1.1.3.2 Lipid rafts

Lipid rafts are small microdomains that can form at the cell surface from sphingolipids and sterols[1,31] and are usually identified by their content of cholesterol, saturated fatty acids and sphingolipids. Ceramide can, in the lipid membrane surface, associate with itself to create rafts containing cholesterol and glycosphingolipids which can fuse together to form large signalling macrodomains.[32] Acidic SMase at the plasma membrane can be triggered by external stimuli leading to the formation of ceramide-enriched areas of the membrane. Specific proteins with an ability to change both composition and size according to stimuli, are incorporated into these rafts. This supports certain protein-protein interactions and can lead to signalling cascades being initiated[33] and thus lipid rafts are believed to play an important role in cell signalling as proteins are very concentrated in these areas allowing for signal amplification.
This is, however, a controversial topic as scientists are divided on whether these structures exist in biological systems, as they are difficult to quantify size-wise and in terms of lifespan.[34] However, recent studies have utilised direct imaging to detect lipid domains with sizes in the μm range.[35]

1.1.3.3 Sphingolipids in other biological processes - signalling agents

Simple sphingolipids such as sphingosine, sphingosine-1-phosphate, ceramide and ceramide-1-phosphate have been linked to several cellular processes such as

growth, differentiation, motility, proliferation, apoptosis and function characteristic deterioration (senescence).[2,36,22] In these processes, sphingosine-1-phosphate and ceramide have been identified as having opposite effects.
Ceramide has been shown to play an important role in cell apoptosis and thus higher levels of the compound leads to cell death. Intracellular ceramide levels are affected by both extracellular stimuli and signals such as oxidative stress, heat-shock, ionising radiation, nitric oxide and many more. Ceramide levels are believed to increase either due to sphingomyelin hydrolysis degradation by either neutral or acidic sphingomyelinases (nSMases and aSMases) or by *de novo* synthesis.[2] Two different mechanisms have been suggested for its involvement in apoptosis. Ceramide has been suggested to either bind to specific proteins after being formed in the plasma membrane, whereupon it influences their activity, behaving as a secondary messenger. The second suggestion involves its behaviour as a signalling lipid due to its biophysical properties. Sphingosine and ceramide both are cytotoxic and can cause growth arrest whereas sphingosine-1-phosphate can inhibit apoptosis and can stimulate growth.[15] Sphingosine-1-phosphate is known to be involved in tumorigenesis and inflammation but also supports the survival of cells.[1]
Even though no confirmation of glycosphingolipid degradation leading to ceramide formation has yet been found the complex glycosphingolipids themselves are vital for gangliosides.[2] The glycosphingolipids play a role in cell physiology and possess functionality to act as antigens, growth factors, binding agents for microbial toxins, cell adhesion mediators as well as being capable of mediating cell signal transduction. Complex glycosphingolipids are important for cell recognition and cell signalling as lipids and proteins in nearby cells can interact with the glycan structures.
It is important to note that the complex and wide range of sphingolipids are all based on the relatively simple and central molecules sphingomyelin and ceramide, formed and degraded by the processes described above.

1.1.3.4 Sphingolipid biosynthesis as a drug target

Sphingolipid metabolism contains many possible drug targets involved in the intricate pathway system. More importantly, this is an ideal way to target pathogens as their sphingolipid biosynthetic pathways often differ significantly from mammals, which has lead to the development of drugs which can target most enzymes involved in this pathway.[37]
The two pathogenic protein targets discussed in Chapter 3 and Chapter 4 are both involved in sphingolipid metabolism and catabolism. $PlcHR_2$ from the opportunistic pathogen, *P. aeruginosa*, is a haemolytic heterodimeric complex possessing phos-

pholipase C and SMase activity. The SMase activity contributes to inhibition of angiogensis and cytotoxicity to endothelial cells.[30] *P. aeruginosa* poses a particular threat to cystic fibrosis patients as they can not clear the bacteria from their pathways, leading to chronic infection. Release of the cytotoxic SMase, $PlcHR_2$, is threatening to patient health and inhibition of the enzyme's SMase activity could potentially lower the number of deaths associated with *P. aeruginosa* infection. The SMase is therefore an extremely desirable drug target.
SPT from *T. gondii*, an apicomplexan protozoan parasite which causes a life-threatening disease in animals and humans (toxoplasmosis), has been chosen as a viable drug target as the enzyme differs sufficiently from that of the host.

1.1.4 Project objectives

The project focusses on two enzymes with key roles in sphingolipid metabolism and pathogen-host interactions. The objective of the first part of the project is to biochemically and biophysically characterise the haemolytic phospholipase C/sphingomyelinase $PlcHR_2$ from the opportunistic pathogen *Pseudomonas aeruginosa*. This important virulence factor has been shown to be one of the key components in the invasion of the pathogen into its mammalian host. The first step includes the cloning, overexpression and characterisation of various constructs of the catalytic subunit PlcH. A biochemical assay is established for the $PlcHR_2$ enzyme. This assay is then utilised to identify new inhibitors as lead compounds for drug development with the help of an in-house chemical library of new substrate analogues. The expression and purification will lead to initial crystallisation experiments laying the foundation of a high-resolution crystal structure analysis of PlcH as well as the $PlcHR_2$ complex.
The second part of the project focusses on the serine palmitoyltransferase (SPT) from the apicomplexan protozoan parasite, *Toxoplasma gondii*. SPT catalyses the first, rate limiting step in the *de novo* biosynthesis of ceramide, the key compound in sphingolipid metabolism. Similar cloning and expression strategies to the ones employed for PlcH are used to express and purify a number SPT constructs. These proteins are characterised biochemically in order to identify the catalytic domain and unravel the catalytic mechanism. Additional biophysical experiments including preliminary small-angle X-ray scattering experiments are performed in order to identify the quaternary arrangement of the protein. Extensive crystallisation experiments are leading the way to a full structure determination by NMR or X-ray crystallography.
The objective of the third part of the project is the crystal structure determination

of a triosephosphate isomerase (TPI) from the hyperthermophilic archaeon, *Sulfolobus solfataricus*. It is crystallised (without and with bound phosphoenol pyruvate (PEP)) and the the 3D structures determined. TPI adapts the most common protein fold, the $(\beta\ \alpha)_8$-barrel. This makes it a model enzyme to study for evolutionary purposes. The crystal structure reveals that, as in other hyperthermophilic archaea, *Ss*TPI forms a tetramer in solution, in order to stabilise the enzyme under the extreme conditions it is exposed to. Furthermore, a PEP bound crystal form with PEP in the active site shows conservation of the active site residues when compared with other TPIs from hyperthermophilic organisms.

Chapter 2

Materials and Methods - Experimental procedure

2.1 Materials

All chemicals and materials utilised were purchased from Sigma-Aldrich unless otherwise stated.

2.2 Cloning and small-scale expression procedures at the OPPF

Initial cloning experiments of *Pa*PlcH and *Tg*SPT constructs were carried out at the OPPF (Oxford Protein Production Facility) using HTP methods. For detailed description of procedures refer to the OPPF Standard Protocol for using a ligation-independent cloning method.[38,39]

2.2.1 Polymerase chain reaction (PCR)

All the polymerase chain reactions (PCR) at the OPPF were performed in 96 well microplates.
Polymerase chain reactions using Phusion Flash Mastermix (NEB) were set up on ice following the manufacturer's protocol. Reactions typically contained 25 μL (bringing it to a final concentration of 1X Phusion Flash Mastermix) 2X Phusion Flash Mastermix, 17 μL sterile H_2O, 3 μL (10 μM) of each primer and 2 μL of template

plasmid (provided by Dr Martin Walsh) for each 50 μL reaction. A summary of the plasmids can be found in Appendix A, Table A.1. PCR was typically cycled as listed in Table 2.1 below.

	Time	Temperature/ °C
Initial denaturation	10 secs	98
Denaturation	0 - 1 secs	98
Annealing	5 secs	60
Elongation (repeat 30 times from step number 2)	15 secs/k bp	72
Final elongation	2 mins	72

Table 2.1: Cycling conditions for PCR using Phusion Flash Mastermix. Once the final elongation had taken place, the samples were cooled to 4 °C.

Results were analysed by agarose gel electrophoresis (section 2.2.2).

2.2.2 Agarose gels

1.25% (w/V) agarose gels in 0.5X TBE buffer (5.4 g Tris base, 2.75 g boric acid and 2 mL 0.5 M EDTA dissolved in 1 L MilliQ H_2O) were run at 100 V for 30 minutes. The DNA sample was mixed 5 : 2 (sample : buffer) with 5 x DNA loading buffer (0.25% (w/V) Bromophenol Blue in 30% (V/V) glycerol) prior to loading. HyperLadder^{TM} I (Bioline Ltd.) was used as a DNA marker ladder to estimate DNA sizes. Visualisation of DNA was attained through the presence of SYBR® Safe DNA stain (Invitrogen Life Technologies^{TM}) in the gel and a UV transilluminator.

2.2.3 Purification of the PCR products by AMPure XP Magnetic Bead Purification

The PCR product was mixed thoroughly with 90 μL resuspended AMPure magnetic beads ensuring a homogenous colour. The mixture was incubated for 5 minutes at room temperature (ensuring the binding of any PCR products larger than 100 bps to the beads). The PCR plate was placed on a magnet plate for 5 minutes or until the solution became clear. Primers and unbound PCR product were removed and discarded with the clear solution whilst the plate remained on the magnet.

200 μL 70% C_2H_5OH was added to each well and plates were incubated for 30 seconds at room temperature whereafter the C_2H_5OH was removed and discarded. This washing step was repeated 2 times. The plates were air-dried for 15 minutes to ensure complete evaporation of the C_2H_5OH. The PCR products were isolated by adding 30 μL of elution buffer (EB), (10 mM Tris pH 8.0), to each well and mixing thoroughly (the beads do not have to go into solution for complete elution). 40 μL of the supernatant was transferred to another PCR plate for later use. Another agarose gel was run as described in section 2.2.2.

2.2.4 In-Fusion reactions

10 μL of InFusion cloning reactions were set up, containing 1 μL (100 ng) of appropriate linearised pOPIN vector (a summary of the vectors can be found in Appendix A, Table A.1), 2 μL of purified PCR product (ranging from 10 to 250 ng), made up to final volume of 10 μL with H_2O. Contents were mixed by pipetting, transferred to a PCR plate and sealed. The reactions were run in a thermocycler at 42 °C for 30 minutes. The InFusion reactions were put on ice once the reaction was complete and immediately diluted with 40 μL of TE (10 mM Tris-Cl pH 7.5 and 1 mM EDTA). 3 - 5 μL aliquots of the diluted In-Fusion reactions were transformed into OmniMaxII (phage resistant) chemically competent *E. coli* cells (section 2.2.5).

2.2.5 HTP Transformation of chemically competent cells

Frozen 50 μL aliquots of chemically competent *E. coli* cells were thawed on ice. 3 - 5 μL of purified DNA expression plasmid (or diluted InFusion reaction, section 2.2.4) was added, gently mixed, and left on ice for 30 minutes. Cells were heat-shocked for 45 seconds in a 42 °C water bath and immediately placed back on ice for 2 minutes. 300 μL Power Broth (PB) medium was added to the cells and then cultured at 37 °C for 1 hour (2 hours for InFusion reactions) before plating onto LB agar plates containing required antibiotic. Plates were grown at 37 °C overnight.

2.2.6 Colony selection using Blue/White screening

Following transformation of InFusion reactions, the cells were spread onto Lysogeny Broth (LB) agar plates. Two replicates were performed for each experiment. A mixture containing antibiotics (carbenicillin at a final concentration of 50 μg/mL or kanamycin at a final concentration of 35 μg/mL), 20% X-Gal in DMF as stock,

resulting in a final concentration of 0.02%, and IPTG (500 mM stock) to a final concentration of 1 mM was prepared. 20 μL PB/LB was added to half of the 1 mL agar plate as growth medium before spreading 5 μL cells. The other half were supplemented with the appropriate antibiotic, X-Gal and IPTG mixture and 25 μL cells were spread onto those agar plates without further addition of media. The plates were shaken to ensure spreading and were then left to dry for 15 minutes. The plates were incubated overnight at 37 °C. For each of the constructs, two white colonies were picked.

2.2.7 HTP culture and glycerol stock preparation

Two white colonies were picked after Blue/White screening and grown in 1.5 mL PB supplemented with antibiotic in a shaker-incubator at 37 °C overnight. Glycerol stocks were prepared by adding 100 μL of the overnight culture to 100 μL of LB with 30% glycerol. Stocks were stored at −80 °C.

2.2.8 Purification of plasmid DNA

The overnight cultures of *E. coli* (grown from two white colonies, OPPF standard protocol, ensuring positive clones for 95% of PCR products) were centrifuged at 5,000 × g for 3 minutes at room temperature and the pellets were processed using a commercial plasmid miniprep kit (Qiagen QIAprep® Spin Miniprep Kit) according to the manufacturer's protocol.

2.2.9 Verification PCR

After extracting DNA following plasmid purification by miniprep, two clones of each construct were verified by a PCR screen. Positive colonies were identified by PCR as described below.

Reactions using KOD Xtreme^{TM} Hot Start DNA Polymerase (Novagen) Master Mix were set up on ice. The KOD Polymerase Master Mix was made in bulk for 52 reactions of 25 μL each (130 μL 10 X KOD buffer, 130 μL 10 x dNTP mix, 52 μL 25 mM $MgCl_2$, 11 μL KOD (2.5 U/μL), 814 μL MilliQ H_2O and 7 μL pOPIN forward primer at 100 μL. Reactions typically contained 22 μL Master Mix, 1.5 μL of 10 μM reverse primer and 1.5 μL of the appropriate construct plasmid. PCR was typically cycled as listed in Table 2.2 below.

	Time	Temperature/ °C
Initial denaturation	2 mins	94
Denaturation	30 secs	94
Annealing	30 secs	60
Elongation (repeat 30 times)	1 mins/k bp	68
Final elongation	4 mins	68

Table 2.2: Cycling conditions for PCR using KOD Xtreme^{TM} Hot Start DNA Polymerase. Once the final elongation had taken place, the samples were cooled to 4 °C.

Results were analysed by agarose gel electrophoresis (section 2.2.2).

2.2.10 Small-scale expression trial

Expression screens were set up using two strains of *E. coli* (B834 (DE3) and Rosetta I (DE3) LysS) and two different induction media (IPTG induction medium ($[IPTG]_{final}$ = 0.5 mM, 20 °C overnight), and Overnight Express^{TM} Instant TB Medium (autoinduction,[40] 25 °C overnight)) for each PCR product in 24 deep-well plates. Appropriate antibiotics were added. Transformation was carried out according to section 2.2.5. 1.5 mL of culture was centrifuged at 5,000 × g for 15 minutes and pellets frozen at −80 °C (for at least 20 minutes before robotic Ni^{2+}-NTA screening (section 2.2.10.1)) to ensure efficient lysis.

2.2.10.1 Ni^{+2}- NTA Miniature Expression Screen

Cells were completely resuspended in 210 μL Lysis Buffer (NPI-10-Tween) (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, 1% (V/V) Tween 20, pH 8.0 (adjusted using NaOH), filtered and stored at 4 °C) with 1 mg/mL lysozyme to Buffer NPI-10-Tween) and 400 units/mL of DNAse Type I. This was left for 30 minutes before centrifuging at 5,000 × g for 30 minutes at 4 °C. The supernatant was put over Ni^{+2}-NTA magnetic beads and purified using a Qiagen Ni^{+2}-NTA Superflow 96 BioRobot Kit which was slightly modified.

2.2.10.2 Sodium Dodecyl Sulphate Poly-Acrylamide Electrophoresis (SDS-PAGE)

0.75 mm precast polyacrylamide gels (Invitrogen) were used. Samples were mixed 1 : 1 with loading buffer (100 mM Tris-HCl pH 6.8, 4% (w/V) SDS, 0.2% (w/V) bromophenol blue and 20% (V/V) glycerol) incubated at 95 °C for 3 minutes to denature the protein and run at 200 V for 30 minutes. Protein sizes were estimated relative to a protein marker ladder (Wide Range (MW 6,500-200,000 Da) SigmaMarker^{TM}). Gels were stained with Instant Blue (Expedeon) dye, left for 30 minutes and destained by covering with water for 10 minutes.

2.3 Protein expression and purification preparation

Whereas initial expression trials were performed at the OPPF, the subsequent scale-up experiments were performed in the Durham University Department of Chemistry.

2.3.1 Preparation of chemically competent cells

Cells of the required *E. coli* strain (usually Rosetta II (DE3) pLysS) were streaked onto LB agar plates containing the appropriate antibiotics (chloramphenicol at 17 μg/mL), and incubated at 37 °C overnight. A single colony was transferred to 5 mL LB medium (with appropriate antibiotics added; ampicillin for DH5α and chloramphenicol for Rosetta II (DE3) pLysS) and the culture was incubated at 37 °C overnight, 200 rpm. In a 100 mL sterile flask, 0.5 mL of the overnight culture was added to 50 mL of inoculated LB media. The new culture was incubated at 37 °C, 200 rpm, until an OD_{600} = 0.5 was reached. The culture was left on ice for 30 minutes and then centrifuged at 3,350 × g at 4 °C for 10 minutes in a sterile 50 mL Falcon tube. All the supernatant was discarded and 50 mL 80 mM $MgCl_2$, 20 mM $CaCl_2$ (ice cold) was added to the pellet. The pellet was resuspended by shaking, then left on ice for 20 minutes. The suspension was centrifuged at 3,350 × g at 4 °C for 10 minutes, the supernatant was discarded and the remaining pellet resuspended in 2 mL 10 mM $CaCl_2$, 10% glycerol (ice cold). The cell aliquots were stored at −80 °C.

2.3.1.1 Transformation of chemically competent cells

200 μL of competent cells were thawed on ice and 2 μL of plasmid, containing the gene of interest, was added, gently mixed and left to incubate on ice for 20 - 30 minutes. Cells were heat-shocked at 42 °C for 45 seconds and replaced on ice. 800 μL of LB media was added to the cells within 2 minutes of heat-shocking and cultured at 37 °C for 1 hour, shaking at 200 rpm, before plating approximately 100 μL onto LB agar plates supplemented with the correct antibiotic. Plates were incubated at 37 °C overnight.

2.3.2 Plasmid preparation

Transformed DH5α cells containing the plasmid of interest were plated onto LB agar plates supplemented with 50 μg/mL of ampicillin and incubated at 37 °C overnight. A single colony was transferred to a 15 mL Falcon tube containing 3 mL of LB medium, with 50 μg/mL ampicillin, and incubated at 37 °C overnight, shaking at 200 rpm.) A 5 mL overnight culture of DH5α cells was centrifuged at 5,000 $\times$ g for 10 minutes and the pellet was processed using commercially available miniprep kits (Plasmid Miniprep Kit No. 2, Biomega Inc.) according to the manufaturer's protocol. Plasmids were stored at -20 °C.

2.3.3 Agarose gels

1.25% (w/V) agarose gels were prepared in 1X TAE buffer (242 g Tris, 57.1 g acetic acid and 18.6 g EDTA in 900 mL dH_2O). 0.01% (V/V) of ethidium bromide (DNA stain) was added to the agarose gel mix. The gel was run in TAE buffer. Each lane was loaded with 6 μL of DNA sample mixed with loading buffer (0.25% (w/V) bromophenol blue, 30% (V/V) glycerol) in a 5 : 1 ratio (5 μL of genetic material and 1 μL loading buffer). A molecular weight marker (Fermentas 1 kb ladder) was used as a reference. Gels were run at 100 V for 1 hour and visualised using a UV transilluminator.

2.3.4 Restriction digest of plasmids

2 μL of pOPINF (for PlcH) and pOPINS3C (for SPT) plasmid containing the desired truncation gene were mixed with 1.5 μL E1 buffer (Promega) and 1 μL HindIII enzyme (Promega). The total volume was made up to 15 μL with dH_2O

whereafter the reaction was incubated at 37 °C for 2 hours and then run on an agarose gel to verify the plasmid size (section 2.3.3).

2.3.5 Colony PCR

Colony PCR was used to check that a specific colony of *E. coli* on an LB agar plate contained the desired gene (expected size) for protein expression.
9 μL of dH_2O was mixed with 0.5 μL of forward primer and 0.5 μL reverse primer (stock concentrations approximately 10 μM (from OPPF), giving final concentrations of 0.25 μM) in a PCR tube (primers corresponding to the individual gene truncations) and a single colony was added. 10 μL of Taq polymerase in buffer was added to the reaction. In a TC-312 PCR machine (Techne), 35 PCR cycles were carried out; denaturation at 94 °C for 30 seconds, annealing at 55 °C for 45 seconds and extension at 68 °C for 100 seconds. The products were analysed by agarose gel (see section 2.3.3).

2.3.6 Plasmid sequencing

All plasmids (PlcH and SPT constructs) were sequenced. This was carried out by DBS genomics, Durham University, using the T7 promoter primer and the T7 reverse primer.

2.4 Protein expression

2.4.1 Small-scale protein expression

Small-scale expression screens were carried out in 15 mL Falcon tubes containing 2.5 mL PB (containing 1% (w/V) glucose, 0.4% (V/V) glycerol), 50 μg/mL ampicillin, 17 μg/mL chloramphenicol and a single *E. coli* colony (containing the plasmid of interest). This was incubated overnight at 37 °C, 200 rpm. 312.5 μL of the culture was tranferred to 12.5 mL Overnight Expres^{TM} Instant TB Medium (TBonex) or YT medium (used as control), 50 μg/mL ampicillin, 17 μg/mL chloramphenicol, all in a 50 mL Falcon tube. The cultures were shaken at 37 °C at 150 rpm for 5 hours (for TgSPT the culture was left at 37 °C until the OD_{600} reached 0.5) whereafter the temperature was reduced to 25 °C and incubation continued for a further 22 - 24 hours. No IPTG was added to the tubes with YT media in order

to not start induction.

2.4.2 Small-scale initial expression assay

Glycerol stocks of PlcH_1, PlcH_2, PlcH_3 and PlcH_4 all in Rosetta I (DE3) pLysS cells (treated at the OPPF) were streaked onto LB agar plates. The procedure for small-scale expression was followed (section 2.4.1). The cultures were centrifuged at
3,350 × g at 4 °C for 20 minutes. The pellets and supernatant medium were stored separately at −80 °C. The pellets (250 - 270 mg) and media were thawed on ice. The pellets were lysed with 1 mL lysis buffer (50 mM Tris pH 7.5, 500 mM NaCl, 20 mM imidazole, 0.2 % Tween® 20 (V/V)), containing 1 EDTA-free protease inhibitor tablet (Roche) and 50 μL each of DNAse and RNAse (5 mg/mL) per 15 mL buffer, vortexed thouroughly and sonicated in an Ultrawave water bath for 2 × 3 minutes. The lysate was centrifuged at 18,400 × g at 4 °C for 30 minutes to pellet the insoluble fraction.

2.4.3 Large-scale protein expression

20 mL cultures were grown up in 50 mL sterile Falcon tubes containing 20 ml PB medium (containing 1% (w/V) glucose, 0.4% (V/V) glycerol), 50 μg/mL ampicillin, 17 μg/mL chloramphenicol and one colony. These small cultures were shaken overnight at 175 rpm at 37 °C. 500 mL of Overnight Expres^{TM} Instant TB Medium (TBonex, a granulated autoinduction culture medium for high-level protein production in pET and other IPTG-inducible bacterial expression systems) was prepared according to manufacturer's instructions in 2 L baffled glass flasks. 50 μg/mL ampicillin, 17 μg/mL chloramphenicol and 10 mL of the inoculating culture were added. The culture was shaken at 37 °C at 150 rpm for 5 hours (for TgSPT the culture was left at 37 °C until the OD_{600} reached 0.5) whereafter the temperature was reduced to 25 °C and incubation continued for a further 22 - 24 hours (for PlcH constructs, SPT Δ159, SPT Δ176 and SPT Δ180). For SPT Δ143, the temperature was reduced to 15 °C and incubation continued for a further 48 hours to maximise expression. Cells were harvested by centrifuging at 4,500 × g at 4 °C for 30 minutes. Cell pellets were frozen at −80 °C.

2.5 Protein purification

During all steps of the purification, the relevant sample fractions were stored and analysed by SDS-PAGE gel (section 2.5.6).

2.5.1 Bacterial lysis

Frozen bacterial cell pellets were thawed on ice and suspended in lysis buffer (50 mM Tris pH 7.5, 500 mM NaCl, 20 mM imidazole, 0.2% Tween® 20 (V/V)). One EDTA-free protease inhibitor tablet (Roche), 50 mg/mL DNAses and 50 mg/mL RNAses were added per 15 mL lysis buffer. Lysis buffer was utilised in a ratio of 1 : 1.5 (pellet mass (g) : Lysis buffer (mL)). The suspension was vortexed and sonicated (3 × 2.5 minutes, 37%, 2 pulses per second). The insoluble fraction was removed by centrifugation at 42,000 × g at 4 °C for 30 minutes. The supernatant was filtered, using a 0.22 μm syringe-filter to remove cell debris and larger aggregates.

2.5.2 Immobilised Metal Affinity Chromatography (IMAC)

The cell lysate was loaded onto an FPLC (Äkta) and injected onto a His trap FF 5 mL column (GE Healthcare Life Sciences) already equilibrated with wash buffer (for PlcH constructs: 50 mM Tris pH 7.5, 500 mM NaCl, 20 mM imidazole and for SPT constructs: 50 mM Tris pH 7.6, 500 mM NaCl, 20 mM imidazole, 25 μM pyridoxal-5-phosphate (PLP) and 5% glycerol (V/V)). The column was run with a gradient of 0 - 100% of elution buffer (for PlcH constructs: 50 mM Tris pH 7.5, 500 mM NaCl, 1 M imidazole and for SPT constructs: 50 mM Tris pH 7.6, 500 mM NaCl, 1 M imidazole, 25 μM PLP and 5% glycerol (V/V)) at a flowrate of 5 mL/min.

Fractions and column flow-through (while loading sample) were collected for analysis. Fractions containing protein (verified by SDS-PAGE gel, section 2.5.6) were dialysed against dialysis buffer (for PlcH constructs: 10 mM Tris pH 8, 200 mM NaCl and for SPT constructs: 10 mM Tris pH 7.6, 150 mM NaCl, 25 μM, 5% glycerol (V/V)), using Slide-A-Lyzer cassettes of appropriate molecular weight cut-off (Thermo Scientific), dialysis tubing (Fischer Scientific) or dialysis tubes (Generon).

2.5.3 Anion exchange chromatography

The dialysed fractions were concentrated using a vivaspin centrifugal concentrator (Generon) or a spin concentrator (Fischer Scientific or Agilent Technologies) as per the manufacturers' instructions, ensuring appropriate MWCO for PlcH and SPT constructs. Concentrated samples and loaded onto an FPLC (Äkta) and injected onto a 1 mL MonoQ 5/50 GL anion exchange column (GE Healthcare Life Sciences) pre-equilibrated with wash buffer (10 mM Tris pH 8, 100 mM NaCl). The column was run with a gradient of 0-100% of elution buffer (10 mM Tris pH 8, 1 M NaCl) at a flowrate of 1 mL/min. Fractions of 1 mL as well as column flow-through (while loading sample) were collected for analysis. This procedure as only used for PlcH_2 and PlcH_4.
Fractions containing protein (verified by SDS-PAGE gel, section 2.5.6) were dialysed against dialysis buffer (10 mM Tris pH 8, 200 mM NaCl) using Slide-A-Lyzer cassettes of appropriate molecular weight cut-off (Thermo Scientific), dialysis tubing (Fischer Scientific) or dialysis tubes (Generon).

2.5.4 Size exclusion chromatography

The appropriate fractions containing protein were concentrated (see 2.5.3) to approximately 2 mL and loaded onto an FPLC (Äkta) and injected onto a 120 mL HiLoad 16/60 Superdex 200 column (GE Healthcare Life Sciences) pre-equilibrated with buffer (for PlcH constructs: 20 mM Tris pH 8, 200 mM NaCl and for SPT constructs: 20 mM Tris pH 7.5, 200 mM NaCl, 25 μM PLP, 5% glycerol (V/V)) at a flowrate of 1 mL/min. Fractions of 1 mL were collected.

2.5.5 Protease cleavage

In order to assess the activity of Human rhinovirus 3C (HRV 3C) protease (Qiagen), before attempting protein cleavage, three reactions were carried out (it is used to cleave between Gln and Gly in the Leu-Glu-Val-Leu-Phe-Gln-Gly-Pro site).[41] Three Eppendorf tubes were used, each containing 5 μL of HRV 3C protease 10X Buffer (Qiagen), 44 μL dH_2O and 1 μL HRV 3C protease (c_{stock}= 2 units/μL). *Tg*SPT Δ143 (purified on an IMAC column, see 2.5.2) column and dialysed against appropriate dialysis buffer (10 mM Tris pH 7.6, 150 mM NaCl, 25 μM, 5% glycerol (V/V)) was added to the tubes; 10 μg, 50 μg and 100 μg respectively. The reactions were incubated at 4 °C for 18 hours with samples being taken at 1 hour, 3 hours,

5 hours and 18 hours for analysis by SDS-PAGE gel (see 2.5.6). When performing tag-cleavage on large scale, the protein sample (with HRV 3 C protease added) was incubated in Slide-A-Lyzer cassettes (Thermo Scientific), dialysis tubing (Fischer Scientific) or dialysis tubes (Generon) of appropriate molecular weight cut-off, immersed in dialysis buffer (10 mM Tris pH 7.6, 150 mM NaCl, 25 μM, 5% glycerol (V/V)) (this step was performed simultaneously to dialysis) at 4 °C for 18 hours. The HRV 3C protease was used at a concentration of 1 μL per 50 μg of protein ((1 unit per 25 μg of protein). The samples were concentrated and loaded onto an FPLC (Äkta) and injected onto a His trap FF 5 mL column (GE Healthcare Life Sciences) as described in 2.5.2. This step was only performed for SPT constructs.

2.5.6 Sodium Dodecyl Sulphate Poly-Acrylamide Electrophoresis (SDS - PAGE)

(Modifications to section 2.2.10.2) 0.75 to 1 mm polyacrylamide gels (12% (V/V) bis-acrylamide resolving and 5% bisacrylamide stacking) were poured using the Mini-Protean Tetra Electrophoresis System. Samples were mixed 1 : 1 with loading buffer (VWR), denatured at 95 °C for 3 minutes and run at 200 V for 35 minutes in 1X SDS Running buffer (10X SDS Running buffer: 144.0 g glycine, 30.2 g Tris and 10.0 g SDS was dissolved in 1 L of dH_2O). Protein sizes were estimated relative to a protein marker ladder as stated in figures. Gels were stained by incubation for 1 - 12 hours in 20 mL Instant Blue (Expedeon) dye and destained by covering with water for 10 minutes.

2.5.7 Native gel

0.75 - 1 mm non-denaturing polyacrylamide gels (separating gel: 1.5 M Tris-HCl pH 8.8, and 0.5 M Tris-HCl pH 6.8, stacking gel) were poured using the Mini-Protean Tetra Electrophoresis System. Samples were mixed 1 : 1 with 1X loading buffer (5X loading buffer: 0.31 mM Tris-HCl pH 6.8, 0.05% (w/V) bromophenol blue, 50% (V/V) glycerol), denatured at 95 °C for 3 minutes and run at 200 V for 35 minutes in 1 x electrophoresis buffer (10 X electrophoresis buffer: 14.4 g glycine and 3.0 g Tris in 1 L of dH_2O). Protein sizes were estimated relative to a protein marker ladder as stated in figures. Gels were stained by incubation for 1 - 12 hours in 20 mL Instant Blue (Expedeon) dye and destained by covering with water for 10 minutes.

2.5.8 Determination of protein concentration

For a quick determination of protein concentration, the absorbance at 280 nm was measured by Nanodrop (Thermo Scientific). To determine a more accurate protein concentration, UV-Vis spectra of a series of dilutions of the protein of interest were taken using a Cary 50 Bio spectrophotometer (Varian). 6 M guanidinium chloride (a denaturant which unfolds the protein) was added to the protein sample and to the buffer blank, in a 500 μL cuvette. Absorbance at 280 nm was measured. Using the Beer-Lambert law ($A = \epsilon \times c \times l$) and the value for the extinction coefficient (from the Expasy Prot Param tool (http://web.expasy.org/protparam/), the concentration was determined. The values obtained using the above-mentioned methods differed by less than 10%.

2.6 Mass spectrometry

All samples prepared for analysis by mass spectrometry were submitted to the Mass Spectrometry Service, Department of Chemistry, Durham University, unless otherwise stated.

2.6.1 Tryptic digest

A tryptic digest was performed to prepare the cleaved protein, PlcH_4, for MALDI-ToF (matrix assisted laster desorption/ionisation time-of-flight) (Autoflex II ToF/ToF, Bruker Daltonik GmBH) using a linear detector. The protein was made up to a concentration of 50 mg/mL in a 50 mM TrisHCl pH 7.55, 10 mM $CaCl_2$ buffer. To a 20 μL (1 mg) aliquot of protein, 1 μL of trypsin solution (20 μg of trypsin, 2 μL of 1% acetic acid and 78 μL of buffer (as above)) was added. The mixture was incubated at 37 °C for 2 - 3 hours wherafter 0.3 mg of dithiothreitol (DTT) was added to the mixture. The final mixture was incubated and agitated at 37 °C for 1 hour.

2.6.2 Mass spectrometry to determine oligomerisation state of SPT

Samples of purified SPT Δ158 at 4.64 mg were dialysed into two different systems using a Slide-A-Lyzer dialysis cassette. One sample was dialysed into MilliQ

H_2O pH 5.65 with 25 μM PLP added, and 4.64 mg of SPT Δ158 was dialysed into 40 mM ammonium acetate pH 6.99 with 25 μM PLP added overnight at 4 °C. The sample in MilliQ H_2O precipitated overnight but was easily redissolved. The sample in ammonium acetate had become cloudy and was filtered to remove aggregates. 1 mL of 3.57 mg/mL of SPT Δ158 in MilliQ H_2O with 25 μM PLP and 1 mL of 0.75 mg/mL SPT Δ158 in 40 mM ammonium acetate with 25 μM PLP was submitted to the mass spectrometry service at the Department of Chemistry at Durham University.
Samples were analysed LC-ESI-QToF Premier (Waters Ltd.) (liquid chromatography electro spray ionisation) to determine the accurate mass.

2.7 Circular dichroism spectrometry

The protein was diluted in a buffer suitable for taking CD measurements (30 mM Na_2HPO_4, pH 8) utilising Mini Clarification Spin Columns (Generon) and washing through with the same buffer. The protein in CD buffer was diluted further with CD buffer in the cuvette.
A 500 μL cuvette with 1 cm pathlength was used for all CD data collections. Data for PlcH_2 and PlcH_4 samples were collected at a concentration of 58 nM and 119 nM respectively. Data for SPT Δ158, SPT Δ176 and SPT Δ180 samples were collected at 0.18 μM, 0.25 μM and 0.23 μM respectively. A Cary 50 Bio Spectrophotometer (Varian) was used to measure the absorbance of the protein between 200 and 800 nm. The spectra were obtained using a Jasco 710 spectrapolarimeter with a thermostat-controlled cell compartment at 20 °C. Three CD runs were carried out for each protein sample (a buffer blank, protein sample followed by a buffer blank). 8 spectra were averaged for each run.
Data was normalised with the data collected for the buffer as a blank. The machine units (of the collected data), θ (measured in mdeg) were converted to $\Delta\ \epsilon$ (units of $M^{-1}cm^{-1}$), molar circular dichroism. This was done using the following equation;

$$\Delta\epsilon = \theta \frac{0.1 \times MRW}{(l \times c) \times 3298} \tag{2.1}$$

where MRW is the mean residue weight (molar mass divided by number of residues), l is the path length (measured in cm) and c is the concentration (in units of mg/mL). The resulting data analysed using the CDPro software (available at http://lamar.colostate.edu/), specifically utilising the CONTINLL, CDSSTR and SELCON3 programmes.[42] The reference set, base sets 1 and 10, were suggested for the PlcH and the SPT data respectively. The output CD data were then analysed

by plotting the experimental data and calculated data versus the measured wavelengths in order to determine which programme yielded the best analysis.[43, 44, 45]

2.8 Bio-SAXS

Biological small-angle X-ray scattering (Bio SAXS) data on all four SPT constructs, at various concentrations, were collected by M. Grøftehauge on B21 at Diamond Light Source. Data reduction was performed in DAWN by James Doutch (automatically performed on the beamline pipeline).

2.9 Enzymatic activity

2.9.1 Establishment of an *in vitro* enzymatic activity assay

An assay was established in order to compare the sphingomyelinase activity of $PlcHR_2$ to that of *Bacillus cereus* sphingomyelinase (*Bc*SMase), a well-characterised standard for SMase activity. It was further used to screen compounds for inhibitory effects. Experiments were conducted using the Amplex® Red Sphingomyelinase Assay Kit (Invitrogen Molecular Probes^{TM}).[46, 47, 48]
The overall assay reaction mechanism is shown below in Figure 2.1.

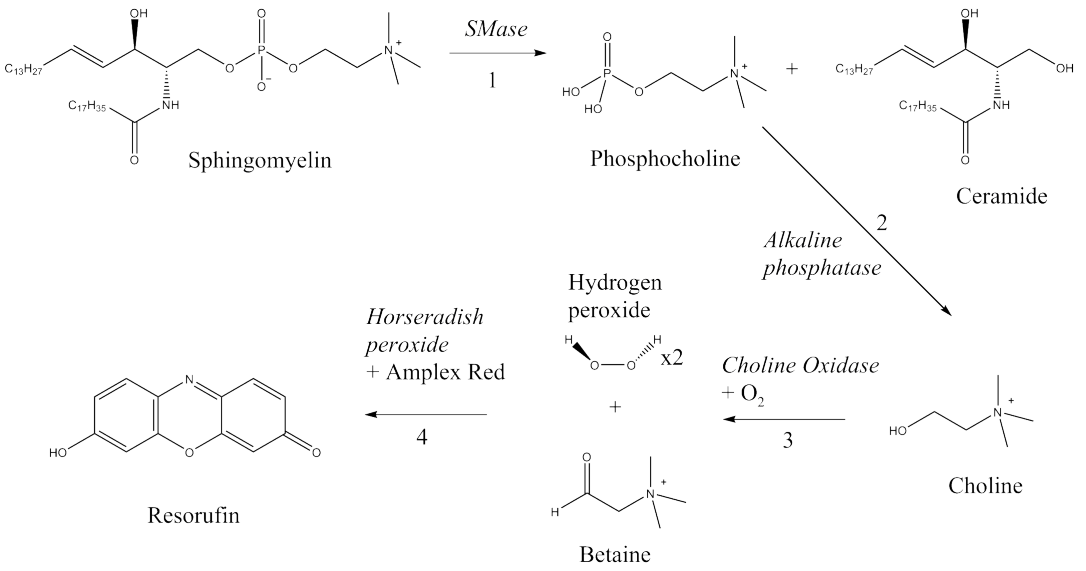


Figure 2.1: Schematic of the Amplex® Red Sphingomyelinase Assay reaction mechanism. Sphingomyelin is hydrolysed by the SMase (1), resulting in phosphocholine and ceramide. The former is hydrolysed by alkaline phosphatase (2), producing choline and phosphate. Choline is oxidised by choline oxidase (3), resulting in betaine and two hydrogen peroxide molecules. Hydrogen peroxide reacts with the Amplex® Red reagent (10-acetyl-3,7-dihydroxyphenoxazine) catalysed by horseradish peroxide (4), producing resorufin, the strongly fluorescent end product.

2.9.2 Experimental protocol for the Amplex® Red Sphingomyelinase Assay Kit

All stock solutions were prepared according to the user manual.[46] Stock solutions included 5X Reaction buffer (0.5 M Tris-HCl pH 7.4, 50 mM $MgCl_2$), functional enzyme stock solutions and substrate stock solution. The following protocol describes the assay for a high throughput reaction, performed in 96-well black (to reduce background signal) microplates (Perkin Elmer®).

One assay was performed in a single well; the equivalent of one reaction. The total reaction volume was set to 200 μL per well.

The sphingomyelinase (*Bc*SMase or the $PlcHR_2$ from *P. aeruginosa*), in storage buffer (50 mM Tris pH 7.4, 50 mM NaCl, 0.1 mM TCEP) at a concentration of 10 mg/mL (0.1 mM), was diluted in 1X Reaction buffer to the required concentration. A volume of 100 μL was utilised per assay (over a range of concentrations). 10 mL of Amplex® Red reagent working solution was prepared by mixing 100 μL Amplex® Red reagent stock solution (10 mM), 100 μL horseradish peroxide stock solution (2 U/mL), 100 μL choline oxidase stock solution (0.2 U/mL), 200 μL alkaline phosphatase stock solution (8 U/mL), 1.0 mL sphingomyelin solution (5.0 mM solution

in 2% Triton X-100) and 8.5 mL 1X Reaction buffer. 100 μL of the enzyme solution was added to each reaction well of the microplate. 100 μL per reaction of 1X reaction buffer was used as a negative control and 100 μL of a 10 μM solution of H_2O_2 was employed as a positive control. To start the reaction, 100 μL of the Amplex® Red working solution (0.1 mM Amplex® Red reagent, 0.02 U/mL horseradish peroxide, 0.002 U/mL choline oxidase, 0.16 U/mL alkaline phosphatase and 0.5 mM sphingomyelin in 2% Triton X-100, all in 1X Reaction buffer) was added to each well resulting in a final reaction volume of 200 μL per well. For assays involving addition of inhibitory compounds, the compounds were added to the wells and mixed with the protein solution before starting the reactions (details are stated in the results section for individual experiments) The reactions were covered to keep out light and incubated at 37 °C for 30 minutes. All fluorescent measurements were made using a BioTek FLx800 microplate reader (using excitation at $\lambda = 540 \pm 17.5$ nm and fluorescence detection at $\lambda = 600 \pm 20$ nm) at a 50% sensitivity. Each reaction was carried out in triplicate unless otherwise stated.
This assay was utilised to test the SMase activity of the PlcH constructs, PlcH_2 and PlcH_4. As a control experiment, purified, circular pOPIN F plasmid (employed as vector for both PlcH constructs) was tested using the same experimental protocol as for the enzymes. No SMase activity was detected (data not shown).

2.9.3 SPT activity assay with [^{14}C]-serine

Labelled [^{14}C]-serine was utilised with palmitoyl-CoA in activity assays to establish the activity of the SPT constructs in crude or purified form.[49,50]
The protein sample was dialysed into 50 mM HEPES pH 7.6, 150 mM KCl, 0.2 mM EDTA, 5% glycerol and 25 μM PLP using a Slide-A-Lyzer dialysis cassette. A typical assay was carried out at a 250 μL volume. The reaction consisted of 20 μM SPT (in 50 mM HEPES pH 7.6, 150 mM KCl, 0.2 mM EDTA, 5% glycerol and 25 μM PLP), 1.6 mM palmitoyl-CoA and 20 μM [^{14}C]-serine. The reaction was incubated at 37 °C for 70 minutes whereafter the reaction was terminated by the addition of 20 μL of 2 : 1 $CHCl_3$: CH_3OH. The solution was thoroughly mixed by vortexing. The organic and aqueous layers were separated by centrifuging at 10,000 × g for 5 minutes. The chloroform layer was transferred to a 1.5 mL non-stick Eppendorf tubes and removed by vacuum centrifugation at 30 °C. The lipid residue was resuspended in 15 μL ($CHCl_3$: CH_3OH) from which approximately 8 μL were applied to a TLC plate in a 40 : 10 : 1, $CHCl_3$: CH_3OH : NH_4OH solvent system. In reactions containing myriocin as inhibitor, 5 μL myriocin dissolved in CH_3OH was added to the reaction to a concentration of 100 μM after addition of

palmitoyl-CoA. After the inhibitor was added, the reaction was incubated at 37 °C for 20 mins before addition of serine ([^{14}C]-labelled or un-labelled). The TLC plate was scanned using an AR-2000 radio TLC Imaging Scanner (Bioscan).

Fluorography was used as an alternative, visual technique. The TLC plate was sprayed with En^{3}Hance Spray (Perkin-Elmer®) and left to dry, twice. A photographic film was placed on the TLC plate, left at −80 °C and typically developed 24 hours later.

2.9.4 Activity assay and product identification by mass spectrometry

SPT Δ158 was dialysed into suitable buffer (50 mM HEPES pH 7.6, 150 mM KCl, 0.2 mM EDTA, 5% glycerol and 25 μM PLP) in a Slide-A-Lyzer dialysis cassette, then transferred to a glass vessel (to avoid plastic-ware polymer contamination). Palmitoyl-CoA and L-serine were added to the vessel at concentrations of 1.6 mM and 20 mM, respectively, to a total volume 250 μL and 20 μM SPT Δ158. The reaction was incubated at 37 °C for 75 minutes. 1.5 mL of 2 : 1 $CHCl_3$: CH_3OH was added and the mixture vortexed. The aqueous and organic layers were separated by centrifugation at 4,000 × g for 5 minutes. The chloroform layer was removed under vacuum at 30 °C, leaving a lipid residue which was resuspended in 500 μL of 2 : 1 $CHCl_3$: CH_3OH and submitted for analysis by mass spectrometry

2.9.4.1 Mass spectrometry of SPT activity assay product

Protein masses were determined using MALDI-ToF (matrix assisted laser desorption/ionisation time-of-flight) (Autoflex II ToF/ToF) using a linear detector. Protein samples at approximately 2 mg/mL in 100 mM NaCl, 10 mM Tris pH 7.5 buffer were analysed on a MALDI ion source. Data were acquired and analysed using FlexAnalysis 2.0 (Bruker Daltonik) by the Durham University, Department of Chemistry Mass Spectrometry Open Access facility.

2.9.5 Thermal Stability Assays

Thermal stability assays were carried out for PlcH_2 and SPT Δ158, using a fluorescent dye. SYPRO Orange dye (Invitrogen) was used as the fluorescent dye as it exhibits a large change in fluorescence upon protein denaturation. It has

an excitation wavelength of 492 nm[51] which reduces the risk of small molecules quenching the fluorescent intensity. The molecule is hydrophobic in nature and thus binds to the hydrophobic protein regions once the protein starts unfolding, and the detected fluorescence levels can be used to judge the degree of denaturation, allowing determination of the melting temperature of proteins.
Protein samples were mixed with 5000X SYPRO Orange dye to a final concentration of 20X. The additives and protein were loaded into a clear plastic 96-well plate, allowing for a broad range of conditions to be tested. To each of the wells, 10 μL additive solution and 10 μL protein-dye solution was added (final dye concentration was 10X SYPRO Orange). Fluorescence at 610 nm was measured in 1 °C increments in a temperature range from $24-95$ °C using a 7500 FAST RT-PCR system (Applied Biosystems). The fluorescent data were plotted against the temperature resulting in a sigmoidal curve of fluorescence as a function of increasing temperature. The inflection point of the curve corresponds to the melting temperature of the protein and was determined by fitting a Boltzmann (sigmoidal) distribution to the curve and then differentiating it, using the QtiPlot 0.9.7.10 software.[52,53,54]

2.9.6 Protein crystallisation

HTP screening was undertaken using a Screenmaker 96+8^{TM} *Xtal* liquid handler from Innovadyne technologies Inc. to produce sitting drops. Two sitting drops of 100 nL reservoir solution + 100 nL protein solution and of 200 nL reservoir solution + 100 nL protein solution respectively, were set up.

For PlcH_4 trays were set up at 1.9 mg/mL using the JCSG+[55] screen (Molecular Dimensions). Manual hanging drop trays were set up with PlcH_2 at a concentration of 1.9 mg/mL with a reservoir volume of 0.5 mL and a drop ratio of 1 : 1. Optimisation screens of the promising conditions identified in initial screens for PlcH_4 were set up using hanging drop trays at a concentration of 2.1 mg/mL with a reservoir volume of 0.5 mL and a drop ratio of 1 : 1.

Initial HTP screens for the SPT constructs were set up with the following screens. JCSG+[55] (Molecular Dimensions), the PACT Premier[55] (Molecular Dimensions), the Index[56] (Hampton), Morpheus[57] (Molecular Dimensions), Clear Strategy 1 and 2 (Molecular Dimensions),[58] Structure Screen I and II[59,60] (Molecular Dimensions) and Midas[61] (Molecular Dimensions). In addition, screens were set up with 100 μM myriocin using the same conditions. Additional screens were set up at

the OPPF using the Hydra^{TM} 96 Robbins Scientific robot (Cartesian, Digilab, microsys). Two sitting drops of 100 nL reservoir solution + 100 nL protein solution and of 200 nL reservoir solution + 100 nL protein solution respectively, were produced. These trays were set up at a concentration of 5.2 mg/mL with the JCSG+[55] (Molecular Dimensions), the PACT Premier[55] (Molecular Dimensions), the Index[56] (Hampton), Morpheus[57] (Molecular Dimensions) and the Wizard 1 and 2 screens. Two identical screens were set up for each screen, one was kept at 4 °C and one at 21 °C. Concentrations of protein were typically between 1.2 and 5.2 mg/mL.

2.9.7 Figures

All figures of chemical structures were created with Chemdraw Std. 13.0 unless otherwise stated.

Chapter 3

The phospholipase/sphingomyelinase $PlcHR_2$ from *Pseudomonas aeruginosa*

3.1 Introduction

Pseudomonas aeruginosa is a Gram-negative bacterium. It is an important opportunistic pathogen which affects immuno-compromised as well as cystic fibrosis patients severely. The pathogen produces various virulence factors aiding bacterial invasion and infection.

In this chapter, my work on the heterodimeric haemolytic phospholipase, $PlcHR_2$, which represents an important virulence factor from the opportunistic pathogen *Pseudomonas aeruginosa*, will be presented. Particular focus is given to the sphingomyelinase activity of this enzyme. After a brief introduction of the bacterium and its various proteins that contribute to pathogenesis, the biochemical and biophysical charaterisation of $PlcHR_2$, and the efforts towards a structural description will be summarised with the ultimate goal of finding novel inhibitors with therapeutic potential.

3.1.1 *Pseudomonas aeruginosa*

Bacteria are a diverse group of prokaryotic (usually unicellular) organisms which have adapted to a wide variety of environments. Bacteria can broadly be divided into two classes, according to the ability of the cells to stain with crystal violet compound also known as pyoctanine (Gram-negative bacteria do not retain this stain whereas Gram-positive bacteria do). Gram-positive bacteria possess a single phospholipid membrane covered by a thick peptidoglycan layer whereas Gram-negative bacteria also possess an extra layer consisting mainly of lipopolysaccharides (LPSs) and proteins as shown schematically in Figure 3.1.

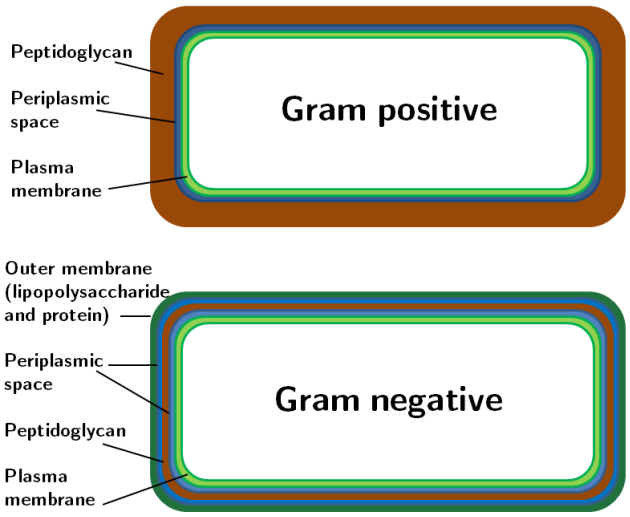


Figure 3.1: Schematic of a Gram-positive and a Gram-negative bacterium.

Pseudomonas aeruginosa is a Gram-negative bacterium. It is part of the family Pseudomondacea and it belongs to the genus *Pseudomonas*, a diverse group of more than 140 species out of which 25 are often associated with humans. Those species linked to diseases are usually opportunistic pathogens such as *P. aeruginosa* and *P. maltophilia*.[62]

P. aeruginosa was first discovered by C. Gessard in 1882[63] due to a distinctive blue pigment, pyocyanine, which is produced by most strains.[62] Its importance as a human pathogen was not discovered until much later. Today its involvement in human disease and mortality is well-established and hence a lot of research has been carried out in order to better understand the pathogen-host relationship. The complete genome of *P. aeruginosa* strain PA01 consisting of approximately 6.3 million base pairs, was thus sequenced by C. K Stover *et al*[64] by genome-shotgun sampling; at the time it was the largest bacterial genome ever to have been sequenced. It was

chosen due to its important role as a human opportunistic pathogen, which is linked to its resistance to most antibiotics and most disinfectants, providing scientist with a unique understanding of the connections between gene complexity, size and ability to adapt to its surroundings.

3.1.1.1 Physical appearance/aspects of *P. aeruginosa*

Pseudomonas aeruginosa is a rod-shaped aerobic bacterium which usually measures 0.5 - 0.8 by 1.5 - 3.0 μm. Most strains move by means of a single polar flagellum,[62] although some strains can possess up to three flagella. In nature it is usually found as a unicellular organism in planktonic form where it can swim using its flagellum but can also exist attached to either a surface or a substrate as a biofilm, depending on the specific strain (section 3.1.4.1). Clinically isolated strains often have surface structures such as pili which are believed to enable the bacteria to better fix at a surface and thus enhance chances of colonisation.[65] Pili are also believed to have antiphagocytic properties (section 3.1.4.1).
The presence of *P. aeruginosa* is often diagnosed by a fruity odour, its ability to grow at elevated temperatures, fluorescence with exposure to UV light (sometimes used to confirm presence in wounds) and its inability to ferment lactose. Isolates of the bacterium can form three distinct types of colonies; small, rough colonies are usually formed by isolates from water or soil whereas clinical samples can produce two smooth colony types - large colonies with flat edges and an elevated appearance (fried egg) or colonies with a mucoid appearance (usually from urinary tract or respiratory secretions) due to alginate slime.[66]

3.1.1.2 Suitable growth environments of *P. aeruginosa*

P. aeruginosa is a ubiquitous bacterium which prefers moist environments. It can be found growing in marshes, soils and coastal marine habitats but it can also grow on surfaces of plants and occasionally on animal tissue.[62,64,67]
The bacterium is non-fermentative aerobe and thus does not acquire energy from carbohydrate fermentation but rather from oxidation.[64] Despite having an aerobic metabolism it can grow under anaerobic conditions if NO_3^- is present as a respiratory electron acceptor.[68] Its oxidative, aerobic metabolism is supported due to the lack of sugar transporters necessary for maintaining glycolytic pathways leading to a dependence on various other carbon compounds.[64] The bacterium's nutritional needs are simple as it can utilise more than 75 organic factors as carbon sources for growth but the simplest medium in which it can grow in the laboratory consists

of acetate and ammonium sulphate (carbon and nitrogen sources). It can even be found growing in distilled water.[62,64] The optimum temperatures for growth are between 25° and 37 °C but unlike other *Pseudomonas* species, *P. aeruginosa* can also grow at temperatures up to 42 °C.
The bacterium has a remarkable tolerance for many types of environment and is able to resist high salt and dye concentrations, weak antiseptics and many commonly used antibiotics (section 3.1.3.3).[64]

3.1.1.3 *P. aeruginosa* - a Gram-negative bacterium with antibiotic resistance

The *P. aeruginosa* bacterium is enveloped by three layers consisting of an inner cytoplasmic membrane, a thin peptidoglycan layer and an outer membrane consisting mainly of lipopolysaccharide, phospholipids and outer membrane proteins (OMPs), as shown in Figure 3.2.[64,69] In *P. aeruginosa*, the LPS has been found to be less toxic than for other Gram-negative bacteria; it generally consists of hydroxy fatty acids, heptose and 2-keto-3-deoxyoctonic acid as well as core- and side-chain polysaccharides.[62] LPS is attached to the outer membrane surface by lipid A (a hydrophobic lipid component of LPS) which is linked to a repeating saccharide chain (the outermost part of LPS) via a core oligosaccharide. The side chain is a variable O antigen which extends into the surrounding environment thus enabling cellular interactions. The O antigen specifies the bacterial serotype specificity.[65]

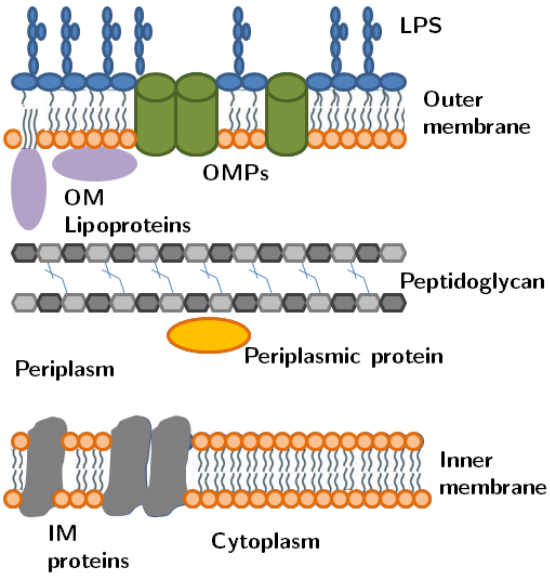


Figure 3.2: Simplified schematic of a Gram-negative bacterium membrane.

Many of the commonly used antibiotics show limited efficacy against *P. aeruginosa* infections as the low permeability of the outer membrane renders them resistant. LPS and outer membrane proteins embedded in the outer membrane protect the bacterium from certain chemicals as it increases the overall negative charge of the membrane (as opposed to Gram-positive bacterial membranes, which just consist of a cytoplasmic membrane and a thick peptidoglycan layer).[64] Many bacteria have evolved sophisticated systems to secrete various proteins including important virulence factors into the surrounding medium. *P. aeruginosa* is found to possess, amongst others, very efficient efflux systems against antibiotics[70] (section 3.1.2) and OMPs have been shown to be actively involved in antibiotic transport, secretion of virulence factors as well as securing the structures which are involved in cell adhesion and motility.[65]
In certain clinically isolated *P. aeruginosa* strains, specific overexpressed OMPs have been linked to resistance, especially against quinolones (a family of antibacterial synthetic drugs).[69]

3.1.2 Secretion pathways in *P. aeruginosa*

Gram-negative bacteria employ up to six types of protein secretion systems in order to secrete a large variety of proteins ranging from ones used for ion uptake to proteins for hydrocarbon hydrolysis.[71,72] These systems are extremely useful for host invasion; they can enable colonisation and can aid in immune response evasion. More importantly, bacteria need a channel through the outer membrane in order to secure energy sources.
Bacteria, like other types of cells, are surrounded by a hydrophobic membrane preventing toxic factors entry, but channels through this membrane allow movement of molecules. Pathogenic bacteria require virulence factors to cross the bacterial cell wall but also the cell plasma membrane of the host. The relevant secretion systems important for *P. aeruginosa* are briefly described below.

3.1.2.1 The Sec and the Tat pathways

As the Gram-negative cell wall consists of an inner and an outer membrane (both hydrophobic) separated by the hydrophilic periplasm, it poses a challenge for large, hydrophilic proteins and toxins being secreted by the pathogen. The protein can either pass from the cytoplasm to the surface of the bacterium in a one-step mechanism or it can briefly stop in the periplasm and then cross the outer membrane in a two-step mechanism.

In the two-step mechanism, the proteins are generally secreted via an export pathway utilised by periplasmic and outer membrane proteins - the most important one of these is the Sec pathway. In order to be secreted via the Sec pathway, proteins must have an N-terminal signal peptide which is cleaved off for the protein to be delivered to the periplasm.[72] Proteins are transported in an unfolded manner. Translocation of proteins in folded conformation through the inner membrane can also occur via the twin arginine translocation (Tat) system, which consists of three components - TatABC.[73,74] It is regulated by pH changes, unlike the Sec system which utilises ATP hydrolysis.[75] Many substrates employing the Tat-system require a cofactor for functionality that they obtain from the cytoplasm. PlcH and PlcN (sections 3.1.7.2 and 3.1.7.3) utilise the Tat system to cross the inner membrane recognised by a specific Tat sequence, the N-terminal signal peptide. This sequence contains two conserved arginine residues. This step is followed by the type II secretions system (T2SS) (the extracellular secretion machinery, Xcp) to cross the outer membrane (as do proteins that cross the inner membrane via the Sec pathway).[75] Tat signal sequences are long (35 amino acids for PlcH) and positively charged (3+ and 6+) which distinguishes them from Sec-dependent signal sequences.

3.1.2.2 The type III secretion system

The type III secretion system (T3SS)[76,77] (also known as the T3SS nanomachine, injectosome) is incredibly useful for pathogens as it enables injection of substrates into the cell cytosol from the pathogen cytoplasm, so long as the two cells are in direct physical contact with each other. *P. aeruginosa* uses this system to inject certain proteins that remain in the membrane, which aid in creating a pore, a translocon, leading from the pathogen into the eukaryotic cells, allowing for other proteins to access the host cells. The T3SS is extremely useful for transporting virulence factors into eukaryotic cells, and thus for acute infections and cytotoxicity.[72] The effectors such as ExoS (a bifunctional type III cytotoxin) and ExoU (an effector protein, cytotoxin and broad-specificity phospholipase A) (section 3.1.6.1) passing through the T3SS are in an unfolded state whereafter they bind to a cognate chaperone before being secreted.

3.1.3 The pathogenicity of *P. aeruginosa*

P. aeruginosa is an important human pathogen that exploits weaknesses in host defences to cause infections and disease. It can infect almost any human tissue

if the immune system is compromised or circumvented in any way. As an opportunistic pathogen, *P. aeruginosa* poses a particular threat to immuno-compromised patients such as those undergoing chemotherapy treatment for cancer and patients co-infected with AIDS (acquired immunodeficiency syndrome).[78] Furthermore, *P. aeruginosa* causes dangerous infections in patients with burns and large open wounds where fatality rates are near 50%.

However the pathogen rarely causes disease in healthy individuals, as the host defence system is generally very effective[79] and human polymorphonuclear leukocyte (PMN)-dependent killing ensures resistance.[80]

Bacterial infections by *P. aeruginosa* are of particular importance in hospitals (nosocomical infections) where they are a significant source of septicaemia in urinary-tract infections especially in catheterized patients and in hospital-acquired pneumonia in patients using respirators or ventilators.[64,79,80] During septicaemia the pathogen often causes thrombosis and vasculitis. Dermatitis, soft tissue infections, osteochondritis, endocarditis, gastrointestinal infections and a variety of systemic infections are all a results of infection with the bacterium.[64] It has been shown that *P. aeruginosa* infections are considerably more dangerous for patients infected with strains that contain a type III secretion system and express the virulence factor, ExoU (ventilator-associated pneumonia)[81] (section 3.1.5). Suppression of the immune response can occur when ExoU impedes the phagocytic cells from eradicating bacterial infections of the lungs. This underlines the importance of the secretion system discussed previously.

The human host defence includes neutrophil elastase (NE), which is secreted during inflammation by both macrophages and neutrophils, is linked to tissue damage occurring during *P. aeruginosa* infection. Mice that lack NE are less likely to suppress a *P. aeruginosa* infection than wild-type mice.[80] Both these results show that NE is a part of a natural defence mechanism.[82] Whilst NE can kill *P. aeruginosa* bacteria, there are many defence mechanisms in place in the bacteria which prevent easy clearance of an infection by the hosts. Examples of these are polysaccharide alginate which is known to possess antiphagocytic properties and rhamnolipids which can bring about necrotic death of PMNs. More rhamnolipids are produced in *P. aeruginosa* biofilms if PMNs are present, thus encouraging biofilm formation providing protection for the pathogen.

3.1.3.1 Cystic fibrosis patients

Probably the most important patient group severely affected by *P. aeruginosa* are people suffering from the genetic condition cystic fibrosis (CF). Cystic fibrosis (with a prevalence of 1:2000 in Caucasians) is a disorder inherited through recessive genes, and often patients die young; the mean lifetime is around 35 but can be increased to more than 50 years when treatment is available.[83] It is a condition that affects a wide range of tissues and organs. Important functions are affected by this disease such as pancreatic function, the liver, sweat glands, small intestines and the airways. It also results in diabetes mellitus, lowered liver function, salt loss and more mucous production hindering the airways.[83] CF sufferers carry a mutation in the *cftr* (cystic fibrosis transmembrane regulator) gene,[84] which is located on chromosome 7.[83] This encodes the CFTR, which is an ABC-type ion channel mainly responsible for chloride (Cl^-) transport across membranes. Mutations in the *cftr*-gene lead to a variety of physiological defects; the mutations cause the CFTR to be faulty, which results in less paraciliary fluid (a non-viscous fluid produced by epithelial cells lining airways during active ion transport) in the lower part of the respiratory tract hindering this non-inflammatory defence mechanism from clearing microbes from the airways. Instead the inflammatory defence mechanism is triggered, involving antibodies and PMNs leading to chronic respiratory inflammations and abnormal lung functionalities.[62] With respect to *P. aeruginosa* infections it is remarkable that neutrophils of healthy humans secrete less NE than those of CF patients,[80] leading to further tissue damage. The most notable effects are in the sinus and lungs with abnormal epithelial cells of the airways (usually obstructed by thick mucous in CF patients) allowing long-term colonisation of these locations by a number of opportunistic pathogens including *P. aeruginosa*. CF patients tend to have severe tissue damage and fibrosed lungs but no worse than seen in patients with other problems.[84] The mucous-filled environment is ideal for certain mucoid strains of *P. aeruginosa* that grow in biofilm.[83] They have been shown to produce more extracellular polysaccharide than other strains. The mucoid strains are often found to produce less toxin A and elastase and in some cases lack a flagella. Most strains isolated from CF patients have been found to contain little or no O antigen, the long polysaccharide side chain of LPS.[62]

Morbidity and mortality in CF patients is mainly caused by *P. aeruginosa* in 90% of cases. It can also result from infection with a *Burkholderia cepacia* complex which cause respiratory failure leading to a necessity for lung transplants or death in severe cases, or by *Staphylococcus aureus*, *Mycobacterium* and several others pathogens. Studies of adult CF patients have shown that more than 80% are infected with *P. aeruginosa* and in the past about 50% of those would die from chronic infection.

New, intensive, pre-emptive therapies mean that fewer children contract chronic infection but eventually all CF patients acquire chronic infections which can not be eradicated.[85]

3.1.3.2 Effect on the cystic fibrosis transmembrane regulator by sphingomyelinases

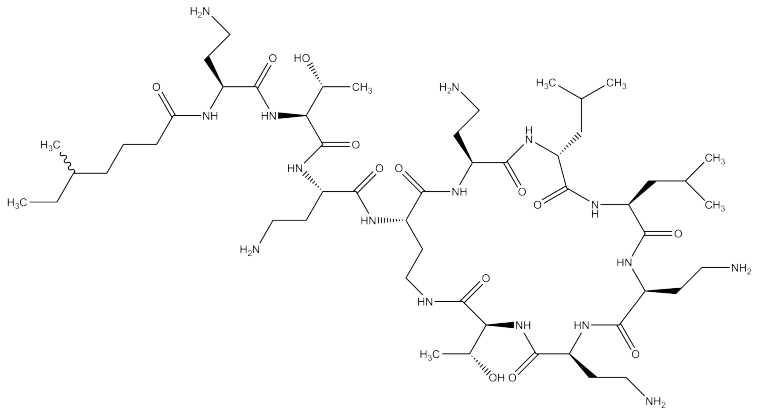
Recently, a link has been discovered between the CFTR and sphingomyelinases (SMases). CFTR function is inhibited by SMases from human respiratory pathogens as the phosphorylation of the regulatory domain required to activate the CFTR is reduced when sphingomyelin is being hydrolysed by SMases.[84] Channel activation occurs by phosphorylation by the cAMP-dependent protein kinase A (PKA) of the regulatory domain, which contains several phosphorylation sites. This results in activation increasing as more sites are phosphorylated but only to a certain threshold. SMase D (from *Corynebacterium pseudotuberculosis* which is associated with repiratory problems) and SMase C (from either *Bacillus anthtracis* or *Staphylococcus aureus*) hydrolyse sphingomyelin to choline and ceramide-1-phosphate or phosphocholine and ceramide respectively. Both larger products have two long hydrophobic chains and thus prefer to remain in the membrane and it has been suggested that the two small components, either choline or phosphocholine, are sufficiently soluble to dissipate after hydrolysis. Neither choline nor phosphocholine are thought to act as inhibitors whereas the longer products, ceramide-1-phosphate or ceramide, are thought to attach to the CFTR, causing inhibition of the Cl^- channel.[86]
In addition, it has been suggested that the CFTR defects increase the chances of bacterial infections in CF patients, thus worsening the chances of pathogenic effects. The CFTR current decreases as SMase C and SMase D inhibits it, causing increased mucous production which supports bacterial colonisation and growth. Tissue necrosis has been linked to SMase D activity and both ceramide-1-phosphate and ceramide are known triggers of inflammation and cell death.[2] These results suggest that targetting bacterial virulence factors such as phospholipases and sphingomyelinases are a possible way of changing the life span and quality of CF patients.

3.1.3.3 *P. aeruginosa* infections treatment

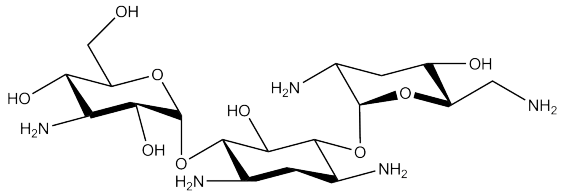
P. aeruginosa infections have not yet been and can not be completely eradicated from CF patients' lungs as the bacterium has multiple adaptive mechanisms; the most important one being its ability to form a biofilm (technique also adapted

by other bacteria) which provides further protection against antibiotics and the inflammatory response.
Chronic suppressive therapy entails early treatment in the form of either prophylaxis or therapy with antibiotics in order to hinder biofilm formation. Lung function has been reported in clinical trials to improve after treatment with combination of already existing antibiotics.[83] Popular antibiotics used to treat CF patient lung infections are colistin (systemic administration) and nebulised tobramycin (administered by inhalation) (Figure 3.3). Tests, both *in vitro* and *in vivo*, have shown that combination therapies involving both oral and intravenous administration is necessary in order to reach high enough antibiotic concentrations in all areas of the lungs.[87] Fluoroquinolones, gentamicin and imipenem (Figure 3.3) are some of the few antibiotics that are effective against *P. aeruginosa*, though unfortunately only certain strains.[88] A treatment method combining the antibiotics gentamicin and carbenicillin (Figure 3.3) is currently used (although resistant strains are emerging),[89] however these infections can not be eradicated and the consequent inflammation and tissue damage ultimately leads to pulmonary failure and death.

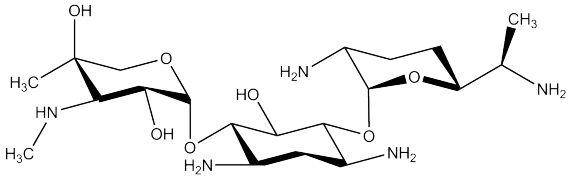
In spite of significant progress in the treatment of *P. aeruginosa* infections, further research is desperately needed in order to identify novel treatments able to eradicate bacterial infections, which could greatly improve the patients' outlook on and quality of life.



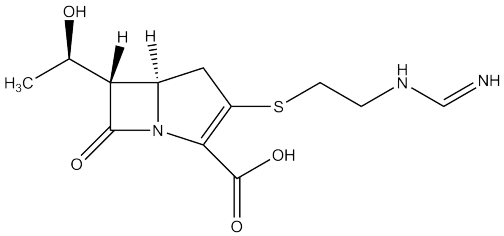
(a)



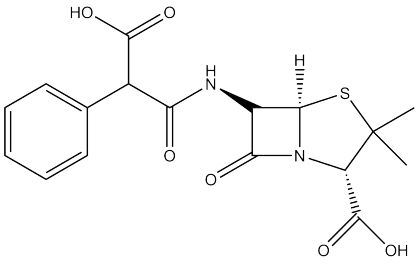
(b)



(c)



(d)



(e)

Figure 3.3: Structures of existing drugs used against *P. aeruginosa* infections. a) Colistin; b) Tobramycin; c) Gentamicin; d) Imipenem and e) Carbenicillin.

3.1.4 Pathogen-host interactions of *P. aeruginosa*

Bacterial infection and colonisation consists of three distinctive invasion stages; attachment and colonisation of the bacteria, local invasion and systemic disease. The bacteria fix themselves to the epithelial cells of the upper respiratory tract using their pili and then link to other cells by interference. Pili adherences[90] in combination with possible overexpression and secretion of a protease enzyme able to degrade fibronectin make colonisation of the tract possible. Fibronectin degradation exposes underlying pili receptors of the tract epithelial cell surface. These receptors appear to be galactose, mannose or sialic acid receptors.[66] Bacterial invasion is a multifaceted mechanism; the most important aids to *P. aeruginosa* invasion and infection are touched upon in the following sections.

3.1.4.1 Biofilm formation and antibiotic resistance

Biofilm formation is a major contributor to wide ranging antibiotic resistance of the bacterium as the drugs are unable to reach the cell when colonies of biofilm form on a surface.[64] Certain *P. aeruginosa* strains produce a mucoid exopolysaccharide[91] which is a repeating polymer of glucoronic mannuronic acid also known as alginate[66] as well as extracellular DNA and proteins which forms the biofilm matrix,[92] anchoring the cells to the surface. This matrix also protects the bacteria from host defences and these mucoid strains are less susceptible to antibiotics than planktonic bacteria.

Biofilm is a major obstacle in treatment of antibiotic resistant CF pathogens leading to combination therapy and further resistance development, fuelling the need for the development of new drugs and antibiotics.[83] Ciprofloxacin (shown below in Figure 3.4) has been used in combination with colistin to target biofilms where the former kills the bacteria growing in biofilms and the latter kills the lower layers.

Furthermore, presumably due to the fact that *P. aeruginosa* evolved to live in various environments in the presence of soil the bacteria have developed a resistance to many naturally occurring antibiotics from moulds and bacili.

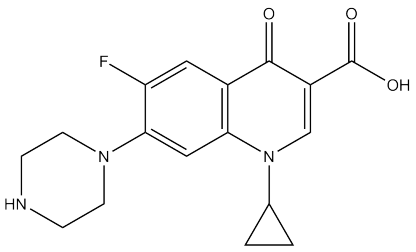


Figure 3.4: Schematic of Ciprofloxacin.

3.1.5 Virulence factors expressed by the bacterium

A large variety of virulence factors are overexpressed and secreted by *P. aeruginosa* during infection, which aid the pathogen in tissue invasion. However, the cellular and molecular mechanism by which these factors work and the nature of the pathogen-host interactions are mostly unknown. The key virulence factors are briefly mentioned below.

3.1.5.1 Extracellular toxins

Extracellular proteases with differing isoelectric points, pH and substrate requirements are found in approximately 90% of *P. aeruginosa* strains.[62] One example is elastase, which breaks down many host proteins including elastin, and alkaline protease, which interferes with fibrin formation, are both present at the pathogen invasive stage and destroy the supporting cellular structures of the host. Some have been linked to haemorrhagic lesion formation leading to necrosis, causing tissue destruction thus facilitating bacterial tissue invasion. Interestingly, LPS which forms part of the outer membrane of the pathogen, is another important virulence factor.[65] The two most studied extracellular protein toxins from *P. aeruginosa* are Exotoxin A and Exoenzyme S.[93] The most toxic extracellular protein is Exotoxin A,[94] a cytotoxin to most eukaryotic cells where it acts as an inhibitor in certain susceptible cells to protein synthesis. It causes ADP ribosylation of the eukaryotic elongation factor 2, causing inhibition of the protein synthesis thus killing the host cell. It is expressed by 90% of the *P. aeruginosa* strains and a dosage of 0.2 μg is enough to kill a mouse host. It has been noted that patients surviving bacterial sepsis have high levels of antitoxin A or low or no amounts of Exotoxin A were produced by the infecting strain. Exoenzyme S is a type III secretion effector which acts as a ADP-ribosyltransferase.[95] It was first detected in blood before the bacteria and is

usually found in bacteria growing in burnt tissue. This may contribute to worsening the function of phagocytic cells to prepare for invasion.

3.1.5.2 Haemolysins and phospholipases

The bacterium also expresses two important haemolysins[96] - lecithinase and phospholipase - which release haemoglobin from red blood cells by destroying them. Additionally, another non-haemolytic cytotoxin is linked to *P. aeruginosa* - it is a pore-forming protein, leukocidin, which is toxic to most eukaryotic cells, neutrophils and lymphopcytes, thus aiding bacterial invasion.[97]
Lecithinase is an α-toxin which possesses haemolytic poperties and is capable of causing myonecrosis, also known as gas gangrene.[98] Last but not least, several phosphatidyl choline phospholipases C (PC-PLCs) are secreted by the bacterium. These include the two phospholipases PlcA and PlcB that are both members of the well-characterised Zn-dependent phospholipase family (section 3.1.7.1). The haemolytic phospholipase C, PlcH, is the most important PLC; it is the focus of this chapter and will be described in section 3.1.7.2

3.1.6 Phospholipases

Phospholipases belong to a class of enzymes that degrade a large variety of phospholipids; they are involved in a number of important processes in eukaryotes such as inflammation and oncogenesis, whereas they are involved in microbial virulence in prokaryotes.[99] Phospholipases can be divided into defined classes each containing sub-groups depending on substrate and cofactor preferences.

Phospholipases are classified according to the position of cleavage site of the phospholipid. These include PLA1, PLA2 and PLB that cleave ester bonds and PLC and PLD which cleave the phosphoester bonds (Figure 3.5). They are further divided according to their enzymatic mechanism and evolutionary relationship.

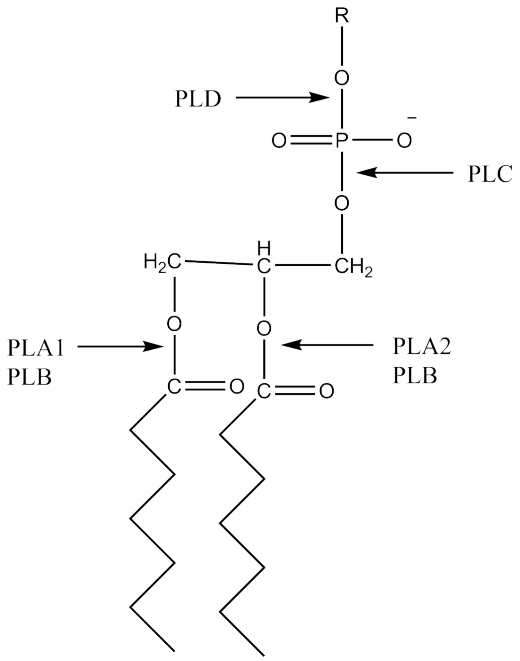


Figure 3.5: Schematic showing the structure of diacyl, a glycerophospholipid, and the cleavage sites for the phospholipases PLA1, PLA2, PLB, PLC and PLD.

3.1.6.1 Phospholipase A and ExoU

PLAs are a diverse group of phoshpolipases with functions spanning from lipid signalling molecule production to playing a role in the structure and function of membranes in eukaryotes. They can be divided into two subclasses, PLA1 and PLA2. The virulence factor ExoU from *P. aeruginosa* possesses phospholipase activity which was discovered as inhibition of its cytotoxicity was linked to eukaryotic PLA2 inhibitors thus proving that is possesses PLA activity. The notion was further supported when it was shown to contain an N-terminal domain which aligns with a domain of patatins, a group of proteins contained in the PLA2 subclass, as well as calcium independent PLA2s.[100] ExoU has been shown to play a role in a number of human infections, in particular hospital acquired pneumonias. Its role in pathogenicity is further supported by the fact that approximately 30% of highly pathogenic *P. aeruginosa* strains produce ExoU.[99]

3.1.6.2 Phospholipase D

Phospholipases D (PLDs) are not frequently produced in prokaryotes but are very common in eukaryotes.[99] PLD produces phosphatidic acid and choline by hydrolysing phosphatidylcholine (PC) or phosphatidylinositol (PI). PLDs belong to

the class PLD Superfamily due to a conserved motif at the active site. The two subclasses of the family consist of plant and bacterial PLDs versus the animal and fungal PLDs. PLDs identified in bacteria are commonly linked to virulence,[101] but little is known about bacterial PLDs. The best studied example of a bacterial PLD is the one from *Y. pestis*, commonly referred to as murine toxin.[102] Cardiolipin synthase is another member of the PLD Superfamily; it forms cardiolipin, a phospholipid that is part of prokaryotic cell membranes and is thus formed in both Gram-positive and -negative bacteria by transfer of a phosphatidyl group from one phosphatidylglycerol (PG) to another. PLDA is an example of a PLD from *P. aeruginosa* and it is located in the periplasm. It prefers PC as substrate, which led to the discovery of the ability of the bacterium to synthesise PC; moreover it is suggested that PLDA is involved in phospholipid signalling events within the bacterium and chronic pulmonary infection.

3.1.7 Phospholipases C from *P. aeruginosa*

PLC activity was first discovered in α-toxin from *Clostridium perfringens*, in 1941 by Macfalane and Knight,[103] and has been observed in many other pathogenic bacteria since. However, the specific role of PLCs in pathogenicity was not discovered until later.

3.1.7.1 The bacterial Zn-dependent PLCs

Zn-dependent enzymes are characterised by three Zn^{2+} ions in the active site that are necessary for catalytic activity. Bacterial PLCs include the important virulence factors α-toxin of *Clostridium perfringens* and PlcB of *Listeria monocytogenes*. They play different roles in pathogenesis. For example the *C. perfringens* α-toxin is cytotoxic to eukaryotic cells and contributes to the severe myonecrosis (gas tissue produced in gangrene). PlcB is vital for the escape of the intracellular pathogen from phagocytic vacuoles into the cytoplasm of macrophages or other cell types infected with the *L. monocytogenes* bacteria.[98]

3.1.7.2 Haemolytic phospholipase C

The haemolytic phospholipase C, PlcH from *P. aeruginosa*, is the founding member of a growing family of phospholipases C, PLCs. These are found in many bacterial and fungal pathogens,[86,104] including proteins from *Bordatella* spp, *Mycobacterium tuberculosis*, *Burkholderia pseudomallei* and *Francisella tularensis*.[86] The first extracellular PLC from *P. aeruginosa* was discovered by Liu[105] in 1966, who noticed its heat-labile haemolytic activity. In the 1980s, PlcH (haemolytic phospholipase C) was found to have phospholipase C and haemolytic activities, and its sequence revealed it as a founding member of a novel class of prokaryotic PLCs.[86,106] This PLC, PlcH, shares no sequence similarity with the previously discovered PLCs and it is not associated with zinc as a cofactor either. PLC from *L. monocytogenes* and *C. perfringens* both possess the same enzymatic activities, but PlcH is approximately twice as large as the *Cf* α-toxin (78 and 48 kDa repectively) and other Zn-dependent PLC enzymes, which may be due to the fact that is is functional as a heterodimer. The importance of PlcH is supported by the fact that it is expressed by 70% of all clinical strains of *P. aeruginosa*.[62]

PLC enzymes have a high cytotoxicity to *E. coli* thus making investigations into related human diseases extremely difficult, however, the full length protein is successfully expressed in its natural host, *P. aeruginosa*.
PlcH has both phospholipase and sphingomyelinase (SMase) activity according to the amount of ceramide present, but it also possesses sphingomyelin synthase (SMS) activity (as shown by Luberto *et al*)[107] where ceramide receives a phosphate headgroup from phosphatidylcholine (PC), producing sphingomyelin (SM) and diacylglycerol (DAG).
Phosphatidylcholine phospholipases hydrolyse PC to give DAG and phospholcholine, but SMase activity hydrolyses SM resulting in ceramide and phosphocholine. Ceramide and DAG are involved in apoptosis, cell differentiation, stress response in eukaryotes and growth regulation and DAG is involved in inflammation, transformation and proliferation. Often increased levels of DAG is used as a diagnostic for mammalian PC-PLC activity.

3.1.7.3 Non-haemolytic PLC

The haemolytic phospholipase C, PlcH, is expressed by *Pseudomonas aeruginosa* alongside a non-haemolytic orthologue, PlcN[108] (section 3.1.7.3). They share a 60% sequence similarity and more than 40% identity; however they have distinctly different functions in infection and disease but are both secreted to the

environment.[86,104] Only PlcH possesses distinct pathogenic properties such as its haemolytic properties whereas PlcN is non-haemolytic. It is noteworthy that both PlcH and PlcN (as well as PLA and PLB)[99] are expressed by all *P. aeruginosa* strains as they are encoded by genes which are connected to more stable parts of the core chromosome. PlcN is expressed when *P. aeruginosa* is grown under Pi-limiting conditions.[109] It is 73 kDa enzyme which is very similar to PlcH with respect to substrate preference. PlcN displays phospholipase C activity and activity on phosphatidylserine (PS) but no evidence is seen for SMase activity.[110] Experiments suggest that PlcH and PlcN consist of distinct domains. PlcN has not been as well characterised as PlcH as no obvious link has been noted with regards to its role in pathogenesis. It has been suggested that PlcN can hydrolyse PS which is exposed during apoptosis of eukaryotic cells (the cytoplasmic facing membrane rich in PS flips to the extra cytoplasmic side). Another suggestion is that it may be involved in *P. aeruginosa* phospholipid biosynthesis alongside PlcH.

3.1.8 The importance of the PlcH family

More than 50 PlcH-homologues have been identified in the Zn-independent phosphoesterase superfamily.[88] To cleave the phosphodiester bond the Zn-dependent and Zn-independent bacterial PLCs employ a different catalytic mechanism supported by a number of biochemical and biomedicinal results as listed below.
Divalent zinc and nickel both inhibit the activity in the PlcH family, whilst the usual inhibitors that work on Zn-dependent enzymes have no effect on this novel class.[86] D609, tricyclodecan-9-ylxanthate, is a popular inhibitor for Bacillus cereus PLC but this does not affect the activity of either PlcH or its homologues from M. tuberculosis.[86]
Calcium can be bound by PlcH; it affects the haemolytic effect of the complex but no change is seen to the enzymatic activity. It is not clear if any cofactor is required for enzymatic activity.

3.1.9 The $PlcHR_2$ complex

PlcH (78412 Da) is co-expressed with an ancillary protein, $PlcR_2$ (17 kDa), forming the heterodimeric complex, $PlcHR_2$. The structural gene coding for PlcHR, *plcH* and $plcR_1$ and $plcR_2$, two overlapping, in-phase genes, make up the *plcHR* three gene operon from *P. aeruginosa*.[88,111]At 3 to *plcH*, the $plcR_{1,2}$ genes are situated and they are expressed from in-phase overlapping genes and have both been shown

to be crucial in the PlcH secretion process. The $PlcR_1$ is synthesized with a Sec-dependent signal sequence and the $plcR_2$ starts at an initiation codon internal to $plcR_1$ (see Figure 3.6)

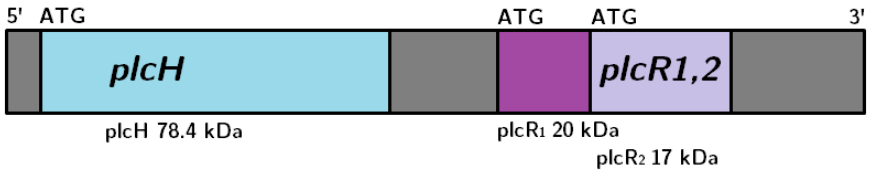


Figure 3.6: The *plcHR* three gene operon showing the overlapping in-phase genes, $plcR_1$ and $plcR_2$, positioned at 3' to *plcH*. ATG marks the initiation codons.

When PlcH is expressed in *P. aeruginosa* under low phosphate conditions, it is associated with a 17 kDa protein when purified but this complex disassociates leaving a single protein of 78 kDa, PlcH. Only the individual protein PlcH is inhibited by tetradecyltrimethylammonium (TDTMA). This suggests that the chaperone protein is not only required for secretion but also offers protection from inhibition and may be involved in the folding of the protein.

The *plcH* gene is expressed without the *plcR* gene in choline media, but is expressed as *plcHR*, which contains a PlcR chaperone protein. PlcH is expressed at high phosphate (Pi) conditions if either choline, carnitine, betaine or dimethylglycine are present or, at low Pi concentrations.[88] $PlcR_1$ and $PlcR_2$ were originally thought to be regulatory proteins but are actually two calcium binding chaperone proteins needed for extracellular PlcH secretion (via the Tat secretion system) and $PlcR_2$ is vital for the haemolytic activity of PlcH.[88]

$PlcR_2$ is a specific chaperone that stays associated with PlcH, forming the heterodimeric complex, $PlcHR_2$. The complex composition has been verified by liquid chromatography electrospray mass spectrometry, showing that the two proteins are at a $1:1$ stoichiometry.[112] $PlcR_1$ consists of 207 amino acids whereas $PlcR_2$ is smaller and only consists of 151 amino acids; the difference is accounted for by a signal sequence only possessed by $PlcR_1$.[111] Without the signal sequence (20 amino acids), the difference is only 36 amino acids, of which 10 are prolines and 15 are various hydrophobic amino acids (such as as Val, Phe, Ile or Ala). When expressed in *E. coli*, $PlcR_1$ is found in the periplasm whereas $PlcR_2$ is in the cytoplasm. Both PlcRs may possess further functions which are still to be elucidated.

3.1.9.1 Sequence analysis of the PlcHR$_2$ components

The sequence of amino acids in PlcH show a multi-domain structure (Figure 3.7) beginning with an N-terminal Tat secretion signal followed by a catalytic domain of about 500 residues. The first 38 amino acids in the 730 amino acid sequence for PlcH are a signal peptide coding for the Tat-signal. The Tat-signal sequence is necessary for PlcH secretion through the inner membrane in *P. aeruginosa*.[75] It is a leader sequence recognised by the export machinery from the cell and can therefore not be part of the active enzyme as it needs to be accessible. This suggests that it is not part of the structural core and that it may not even be folded. The signal sequence is usually hydrophobic and may even be cleaved *in vivo*. There are also two C-terminal domains whose function is unknown. It has been suggested that these domains contain four Ca^{2+}-binding domains (found using structure prediction programmes such as I-TASSER)[113] but their biological role remains yet to be determined (they may be involved in the ability of the enzyme to interact with the cell membrane in order to access substrates).[75] When aligning the sequence of PlcH and the homologues, there is a unique region with 24 amino acids, and this region and the RGD (Arg-Gly-Asp) motif are believed to take part in the interaction with PlcR$_{1,2}$.[86] Residue Thr178 is important for activity and studies have shown that an active site mutant, Thr178Ala, resulted in a significant reduction in activity.[30]

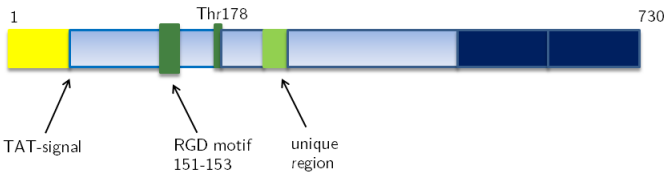


Figure 3.7: The PlcH multi-domain structure highlighting the Tat-signal, the RGD motif, Thr17 and the unique region.

3.1.10 PlcHR$_2$ and its role in pathogenicity

The PlcHR$_2$ is an important enzyme due to its proven biological role in pathogen-host interactions as detailed below.

3.1.10.1 Cytotoxicity of PlcH

$PlcHR_2$ has been shown to be cytotoxic to macrophages and neutrophils.[30] It exhibits haemolytic activity to sheep and human erythrocytes.[112] This effect has been studied on blood agar plates with bacteria expressing both the *plcH* and the *plcR* genes in a T7 expression system in *E. coli* both together and separately, revealing more about their individual roles in haemolysis.[111]

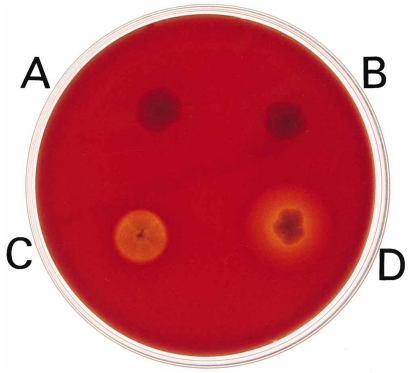


Figure 3.8: Blood agar plate used to demonstrate the haemolytic activity of *plcH* and *plcR* expressed both together and separately in *E. coli* BL21 (DE3) in a *p*GEM vector under the T7 promoter. A) Expression without the *plcH* and *plcR* genes in the *p*GEM; B) expression of *plcR*; C) expression of *plcH* and D) expression of *plcHR*. Reprinted with permission from Figure 1, Cota-Gomez *et al.*[111]

The haemolytic area observed on the blood agar plate is much larger than the area of bacterial growth when both the *plcH* and the *plcR* genes are on the same plasmid under the T7 promoter and are being expressed by *E. coli* BL21(3DE) cells. However, when *plcH* is expressed alone, under the same conditions and from the same system, the haemolytic area is much smaller and only just extends beyond the bacterial growth area. No evidence of haemolysis is seen when *plcR* is expressed alone under the same conditions and from the same system. It was suggested that this either meant that less PlcH was expressed or was unable to be secreted in cells without the *plcR* gene. The former suggestion was discarded as the expression was controlled by the T7 promoter in all cases.[111] PlcH is affected by PlcR posttranslationally, as they form a heterodimeric complex, and so PlcH secretion is not as efficient when lacking the *plcR* gene, even though no effect is seen on expression levels.

3.1.10.2 Membrane vesicles

Membrane vesicles (MVs) are structures shed from *P. aeruginosa* bacteria that contain LPS; they appear during the logarithmic growth and more are produced when small amounts of antibiotics such as gentamicin are present.[114] Active enzymes are present in the MVs, amongst these PLC, and it has been suggested that MVs may be used to deliver virulence factors. Studies with wildtype and knockout mutant, PAO1WT and PAO1ΔR, show different amounts of PLC activity in the corresponding MVs which may indicate that *plcR* is useful in the correct localisation of PlcH within the MVs.

3.1.10.3 Sphingomyelinase activity of $PlcHR_2$

Recent studies have shown that PlcH functions as a sphingomyelinase (SMase)[30] and that the SMase activity of PlcH may be more critical to pathogenesis than its phosphatidylcholine phospholipase C (PC-PLC) activity,[86,112] since the cytotoxocity of PlcH is linked to its SMase activity.[115] The catalytic mechanism has not been elucidated but a proposed mechanism shown below in Figure 3.9 is based on the acid phosphatase, *F. tularensis novicida* AcpA (section 3.1.7.2). One of the active site residues of *Ft*AcpA, Ser196, corresponds to Thr178 in PlcH. In the proposed catalytic mechanism, Thr178 is deprotonated, then activated by an unknown divalent metal ion. A nucleophilic attack by the oxygen atom on Thr178 on the substrate phosphate group leads to an intermediate which remains bound at the active site and the release of ceramide (when the substrate is sphingomyelin), whereafter it is hydrolysed resulting in the formation of phosphocholine (when the substrate is sphingomyelin). Thr178 is believed to be involved in the enzymatic SMase activity, as site directed mutagenesis of this residue results in the loss of SMase activity.[30] PC-PLCs hydrolyse phosphatidylcholine (PC) and produce diacylglycerol (DAG) and phosphocholine. SMases hydrolyse sphingomyelin (SM) and produce ceramide and phosphocholine. Ceramide and DAG are important components of eukaryotic cell membranes and they can often act as antagonists towards each other in mammalian systems.[86] They are involved in various processes such as apoptosis, regulation of growth and differentiation. The latter is affected by ceramide levels whereas DAG has been observed to be involved in inflammation, transformation and proliferation.[28,116] The bacterial $PlcHR_2$ produces this eukaryotic secondary messenger which may be the reason for its cytotoxic effects.

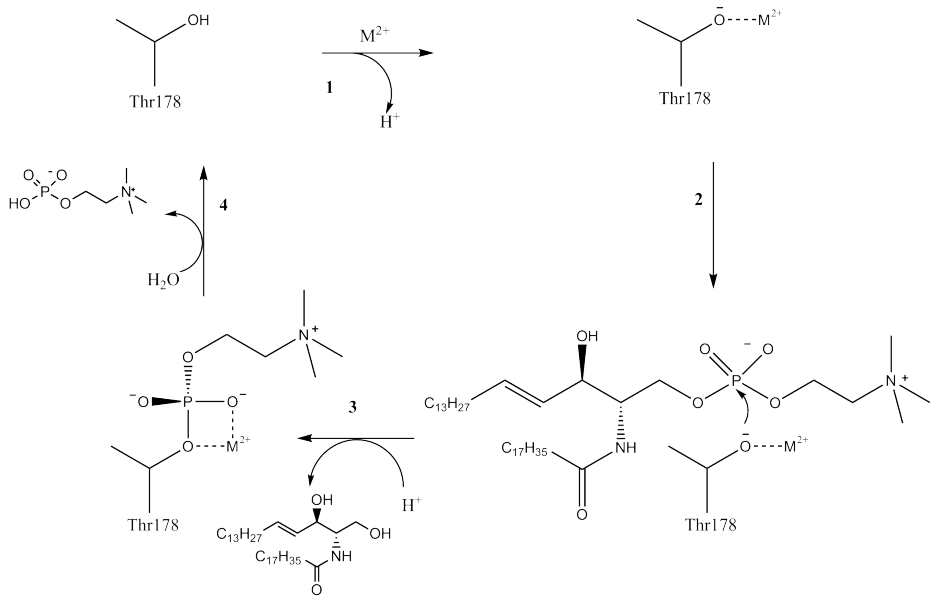


Figure 3.9: Proposed catalytic mechanism of *Pa*PlcH. Thr178 in the active site is deprotonated and then activated by and unknown divalent metal ion. The Thr178 oxygen atom performs a nucleophilic attack on the sphingomyelin phosphate group, forming an intermediate bound to the enzyme and the release of ceramide. The intermediate is then hydrolysed and phosphocholine is released.

3.1.11 Chapter objectives

This section describes the work towards biophysical characterisation of PlcH from *P. aeruginosa* as well as attempts at crystallisation. PlcH is part of a heterodimeric complex, $PlcHR_2$, which has previously been crystallised,[117] however, the structure has not yet been solved. In order to better investigate the structure of $PlcHR_2$, twelve DNA constructs for PlcH were designed, cloned and subsequently tested in expression trials, to aid the identification of the domain structure and produce the minimal catalytic domain for subsequent crystallisation trials. Several biophysical characterisation experiments were conducted with the expressed and purified proteins. In parallel, a commercially available sphingomyelinase activity assay kit suitable for high throughput screening was established for identification of new potential inhibitors of the SMase activity of the wild-type $PlcHR_2$ complex.

3.2 Results

This section describes the work towards a crystal structure determination of PlcH from *P. aeruginosa*, from cloning and expression to a detailed biochemical and biophysical characterisation of the protein.

3.2.1 Cloning and small-scale expression trials at the OPPF

3.2.1.1 PlcH construct design

In order to determine the domain boundaries of PlcH, the sequence was compared to that of a known structure. The haemolytic phospholipase C, $PlcHR_2$, and acid phosphatase, AcpA (respiratory burst inhibiting acid phosphatase), from *F. tularensis novicida*[118] (an intracellular bacterium) share a sequence identity of just under 20% but a number of active sites and residues are conserved. The two enzymes are remotely related, as AcpA is part of a superfamily containing many bacterial enzymes, amongst these, phospholipases C (PLCs). A sequence alignment (Figure 3.10) of AcpA and the significant region of PlcH (residues number 1 to 481), was carried out in Clustal Omega.[119,120] Based on the domain prediction following sequence alignment, the starting points of 12 constructs were determined in order to try and include as many of the essential structural features as possible. The cleavage sites in the sequence were determined by the amino acid sequence as hydrophilic groups may enhance protein solubility.

The sequence alignment highlights the residues which are highly conserved within the superfamily containing AcpA and PlcH. In the crystal structure of AcpA, a vanadate or another metal ion is held in the active site by ten residues which are conserved in enzyme homologues within the superfamily and in PlcH, five of these are conserved.[118] The residues concerned are Glu64, Asn65, His127, His371 and Asp407 (numbers correspond to sequence alignment in Figure 3.10). Despite the low identity between the two sequences overall, the identity within the active site is significantly higher. As well as the five residues mentioned above, further four residues are of importance. Thr178 and Glu358 of PlcH align with Ser196 (a nucleophile) and Asp408 (involved in metal binding) of AcpA, respectively. Another important ion pair are Asp364 and Arg401 which are equivalent to Asp414 and Arg435 of AcpA. The conservation of these nine key residues suggest that the octahedral metal-binding site as well as the hydroxyl nucleophile of AcpA are conserved in PlcH.

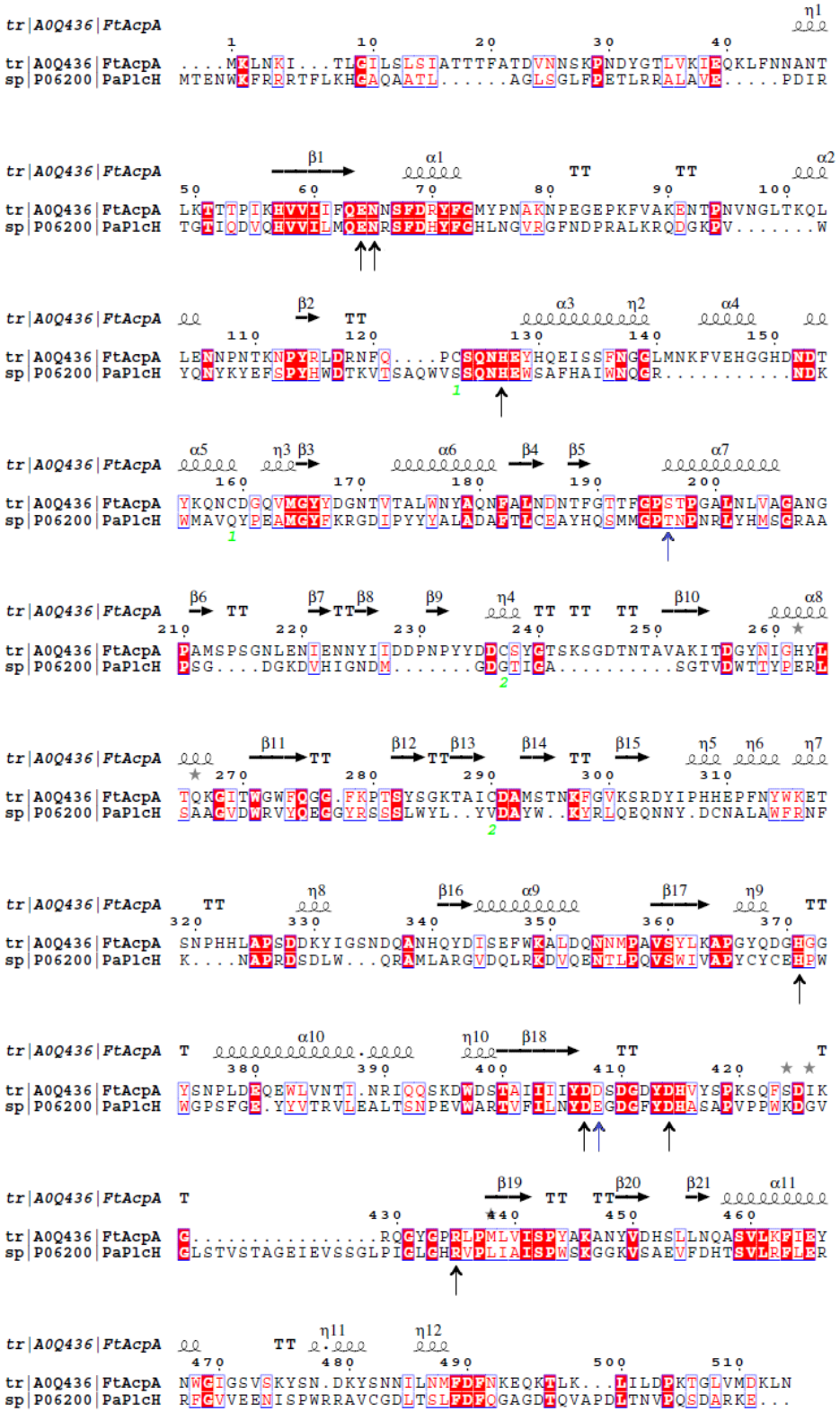


Figure 3.10: Sequence alignment of the *F. tularensis* AcpA sequence and the first 481 residues of *P. aeruginosa* PlcH produced in ESPript 3[121] (available at http://espript.ibcp.fr) with secondary structure overlay of AcpA. Residues enclosed in a red box are identical, residues enclosed in a blue frame with red letters are of similar type and the rest have a low identity score. α denotes α-helices, β denotes β-sheets, TT denotes strict β-turns and η denotes 3_{10}-helices. The vertical arrows highlight the five conserved active site residues and four further important residues.

The constructs decided upon based on the sequence alignment and specific residues are described in Table 3.1 and shown schematically in Figure 3.11.

Construct label	aaN	aaC	pOPIN vector	Tag	PCR product size	MW/ kDa	Plasmid name
1	2	730	E	C-His	2214	82.54	*pPlcH-1*
2	2	730	F	N-His	2214	82.54	*pPlcH-2*
3	40	686	E	C-His	1968	73.54	*pPlcH-3*
4	40	686	F	N-His	1968	73.54	*pPlcH-4*
5	40	699	E	C-His	2007	74.92	*pPlcH-5*
6	40	699	F	N-His	2007	74.92	*pPlcH-6*
7	40	730	E	C-His	2100	78.27	*pPlcH-7*
8	40	730	F	N-His	2100	78.27	*pPlcH-8*
9	40	712	E	C-His	2022	76.42	*pPlcH-9*
10	40	712	F	N-His	2022	76.42	*pPlcH-10*
11	48	473	E	C-His	1305	48.64	*pPlcH-11*
12	48	473	F	N-His	1305	48.64	*pPlcH-12*

Table 3.1: Summary of the 12 PlcH constructs, including construct label, N- and C-terminal, pOPIN vector employed, construct size, molecular weight and plasmid label. The molecular weight represents the protein without the His_6 tag. All proteins contain a 3C cleavage site.

These constructs were all expressed from a pOPINE (C-terminal tag) or pOPINF (N-terminal tag) vector and contain a His_6 tag at the specified terminal. The hexa-histidine enables IMAC purification.

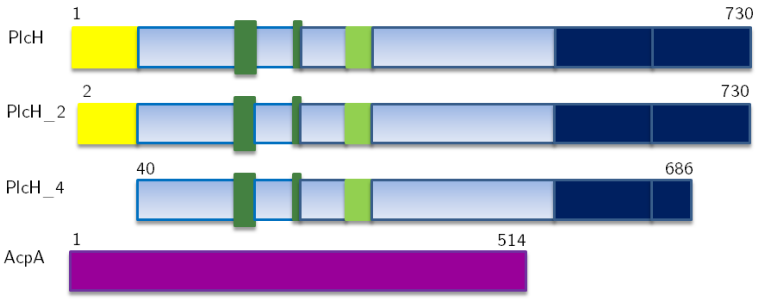


Figure 3.11: Schematic of full length PlcH and two relevant constructs, PlcH_2 and PlcH_4. This is compared to the *F. tularensis* AcpA (shown in purple). Start and end residues are labelled.

Three PCR experiments were carried out before successfully cloning all constructs and all PCR products were analysed by agarose gel. The products from the third and final PCR experiment are shown below in Figure 3.12.

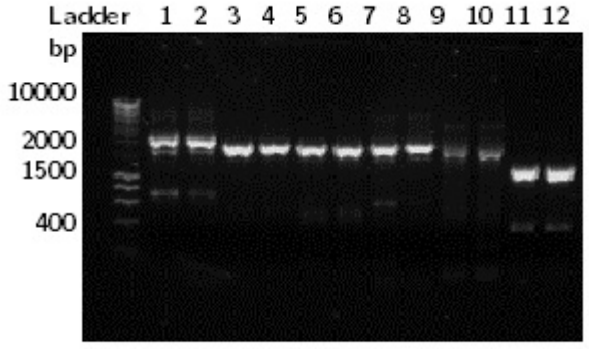


Figure 3.12: 1.25% agarose gel of PCR products, constructs 1 to 12. Product sizes are estimated relative to a Bioline Ltd. Hyperladder^{TM} I.

Figure 3.12 shows that that constructs 1 and 2 (C- and N-His_6 tags) are slightly larger than 2000 bps and the two shortest, 11 and 12 (C- and N-His_6 tagged) are less than 1500 bps. This is consistent with the expected sizes (see Table 3.2.1.1), 2214 and 1305 bps, respectively. The bands representing constructs 9 and 10 (C- and N-His tagged) on the gel are not as strong as for the other PCR products but they are still present at the expected size. The agarose gel confirms that the PCR resulted in the correct size products.

The purified, linearised expression inserts were inserted into the appropriate pOPIN vector (Table 3.1) via an InFusion reaction. The InFusion reaction products were transformed into OmniMaxII cells (section 2.2.4). Following the transformation of the InFusion reaction products, the cells were plated out and a Blue/White screening (section 2.2.6) was performed. Two white colonies (in white colonies, the vector contains the correct gene) were picked for each construct (1 - 12), resulting in two clones (Clone 1 and Clone 2) of each construct. The plasmid DNA for the two clones of each of the 12 constructs was purified by mini-prep.
Mini-preps were performed for all 2 x 12 clones followed by verification PCR experiments in order to establish that the correct plasmids had been obtained (as described in section 2.2.9, Table 3.1) and the products obtained were analysed by agarose gel (section 2.2.2). The gel is shown below in Figure 3.13.

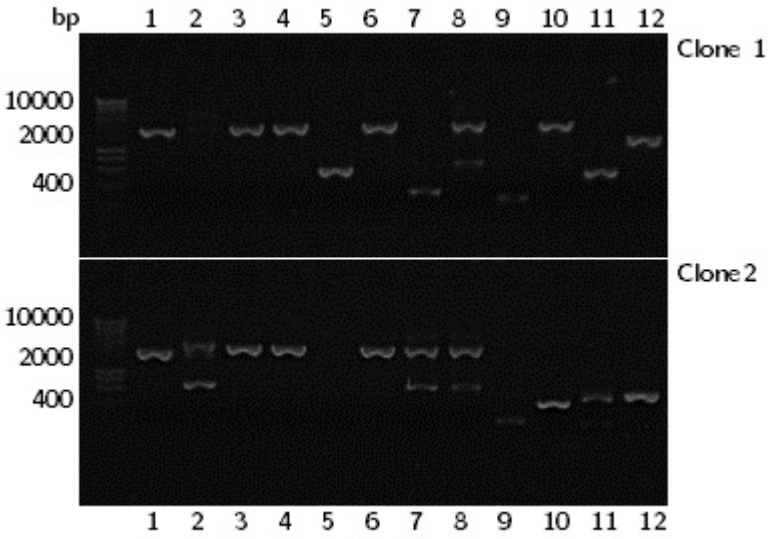


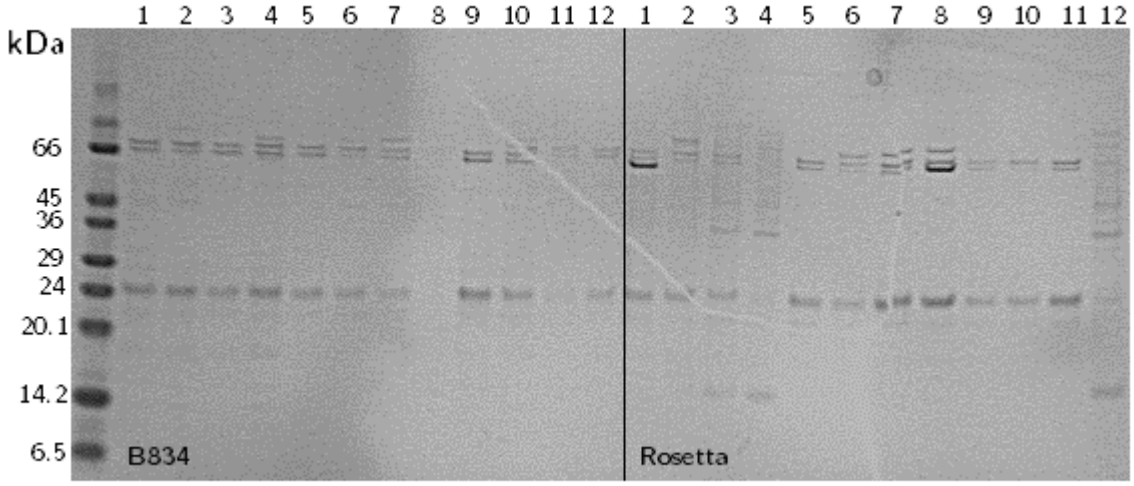
Figure 3.13: 1.25% agarose gel of PCR verification products, constructs 1 to 12. Product sizes are estimated relative to a Bioline Ltd. Hyperladder^{TM} I.

Both clones were derived from cells containing the same construct and thus the bands corresponding to the same construct should be in identical positions for the two clones on the agarose gel. Figure 3.13 clearly shows the PlcH constructs (1 - 12, see Table 3.1), one of each of the clones is closer to the correct plasmid size than the other. The band seen on the gel may also vary in strength between the two clones. Only the best clone for the 12 constructs was taken forward for small-scale expression trials ie. the clone with the correct molecular weight was chosen and if both clones yielded the same results, the one with the strongest band was chosen. For constructs 1, 2, 3, 4, 6, 7 and 8, clone 2 was chosen and for constructs 5 9, 10, 11 and 12, clone 1 was chosen as the most successful clone resulting in the correct plasmid size.

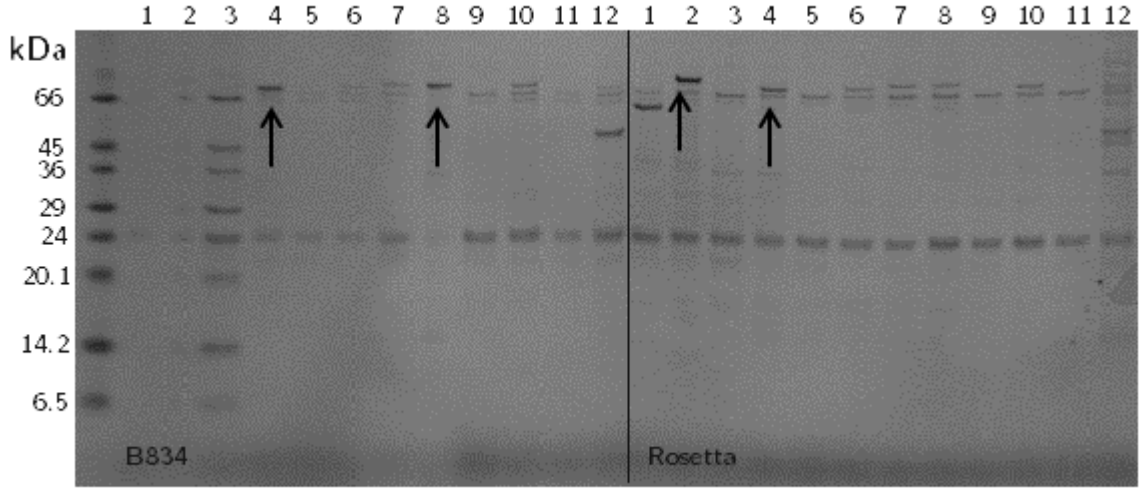
Based on the results obtained from the verification PCR (section 2.2.9) and the chosen clone (1 or 2) for each of the 12 purified constructs (1 - 12), the 12 chosen plasmids were transformed into different *E. coli* strains expressed under different conditions.

Two strains of chemically competent *E. coli* cells were used for transformation and expression trials of the 12 PlcH constructs; B834 (DE3) and Rosetta I (DE3) pLysS cells. Colonies from each strain were picked for all 12 constructs, and each was expressed using different methods of induction; IPTG medium (IPTG induction) and Overnight Express^{TM} Instant TB medium (autoinduction). The expressed cultures

were centrifuged and the supernatant was analysed using a 12% SDS PAGE. The results are shown in Figure 3.14 below.



(a)



(b)

Figure 3.14: 12% SDS PAGE of 5 μL of resolved, purified protein (for constructs PlcH_1 to PlcH_12). Proteins were expressed from two *E. coli* strains; B834 (left side of gels) and Rosetta I (DE3)pLysS. Expression was carried out using IPTG induction medium (Figure (a)) and OvernightExpress^{TM} Instant TB autoinduction medium (Figure (b)). Proteins were stained with Instant Blue (Expedeon) dye and protein size was estimated relative to a Wide Range SigmaMarker^{TM}. Arrows indicate strong bands at the expected size.

Figure 3.14 (a) and 3.14 (b) provide information on the extent of protein ex-

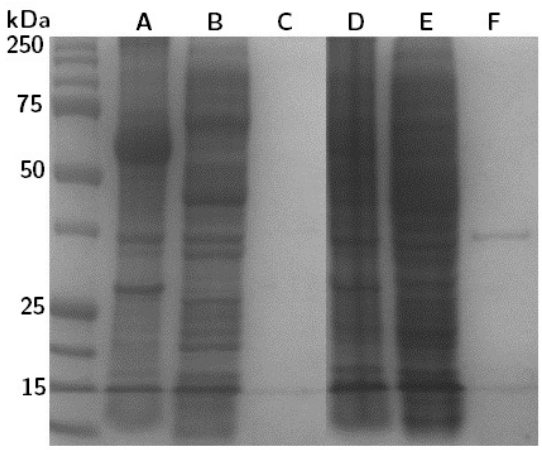
pression under different conditions. No particularly strong bands were visible at the correct molecular weights for the 12 PlcH constructs expressed using IPTG induction as seen in Figure 3.14(a).
In Figure 3.14(b) however, four strong bands are present at the expected molecular weight (see Table 3.1) for the PlcH constructs expressed in the autoinductive medium (see Figure 3.14(a)). The strong bands were seen for constructs PlcH_4 (B834 cells) PlcH_8 (B834 cells), PlcH_2 (Rosetta I (DE3)pLysS cells) and PlcH_4 (Rosetta I (DE3)pLysS cells) all expressed in the OvernightExpress^{TM} Instant TB autoinduction medium. The respective molecular weights are 73.5, 78.3, 82.5 (full length construct) and 73.5 kDa. In general, the PlcH constructs expressed in Rosetta I (DE3) pLysS cells gave rise to stronger bands than those expressed in B834 cells, thus more protein was expressed by Rosetta I (DE3)pLysS cells. Furthermore, larger constructs (higher molecular weight) gave rise to stronger bands (PlcH_2 and PlcH_4). Those constructs with N-His tags also gave stronger bands (better expression) compared to those with C-His tags.
The highest levels of expression were seen for the two N-His tagged proteins, PlcH_2 and PlcH_4, in Rosetta I (DE3) pLysS cells employing the autoinduction method.

3.2.2 Large-scale expression and purification

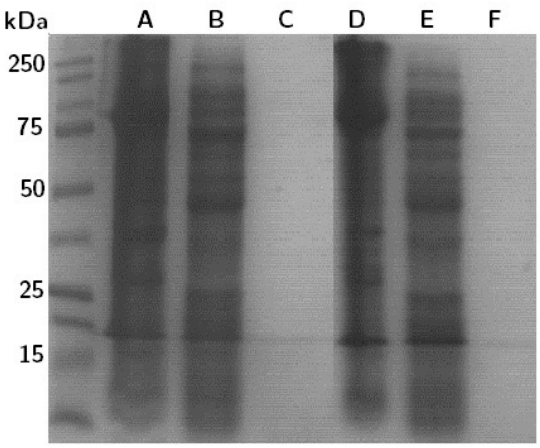
Transformation of the plasmids of constructs PlcH_2 and PlcH_4 into a range of strains of *E. coli* (BL21, BL21 (DE3), BL21 (DE3) Rosetta II, Tunetta and Rosetta II (DE3) pLysS) was attempted but transformation was only possible into BL21 (DE3) cells. Expression levels were below detection levels. Glycerol stocks of Rosetta I (DE3) pLysS and Rosetta II (DE3) pLysS cells containing all the relevant PlcH construct plasmids (obtained from the OPPF) were employed instead. Only overnight cultures from Rosetta I (DE3) pLysS yielded growth.

3.2.2.1 Initial expression assay

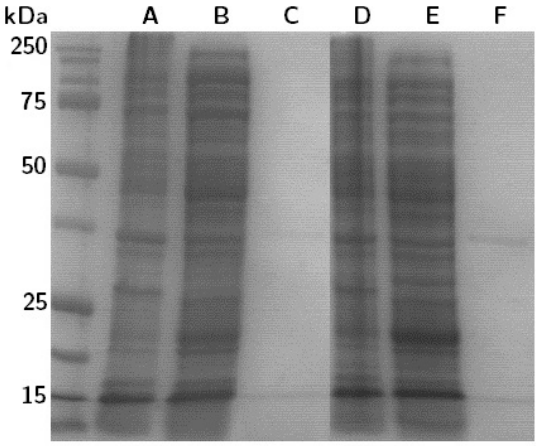
An initial expression assay was carried out to test for leaky expression (protein secretion during expression). A small-scale expression was carried out (section 2.4.2) in autoinduction media (TBonex) and YT media (without addition of IPTG). Cell lysate and media was analysed on a 12% SDS-PAGE gel (section 2.5.6). The resulting gels are shown below in Figure 3.15.



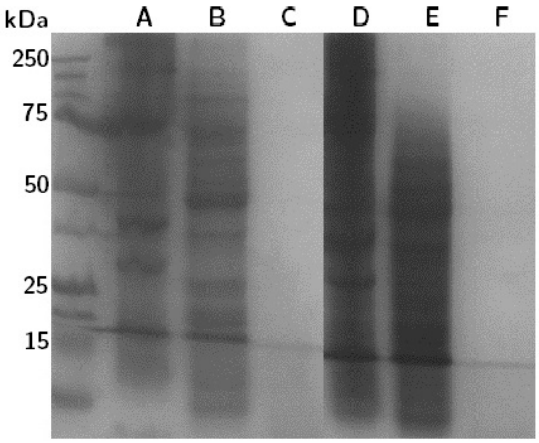
(a)



(b)



(c)



(d)

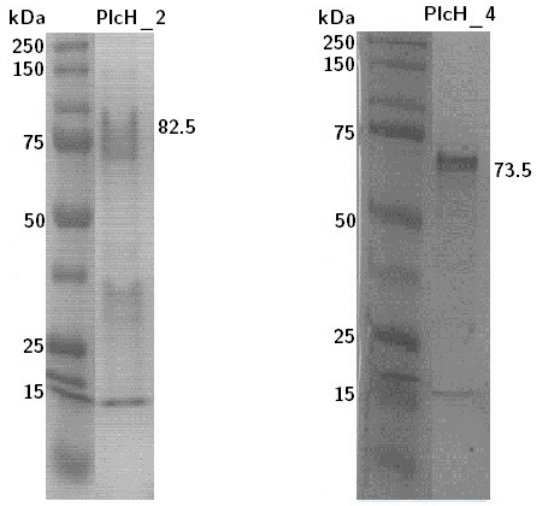
Figure 3.15: 12% SDS PAGE of 10 μL of resolved sample for construct a) PlcH_1, 82 kDa; b) PlcH_2, 82 kDa; c) PlcH_3, 73 kDa and d) PlcH_4, 73 kDa. A) Insoluble fraction (after cell lysis) from expression in TBonex. B) Supernatant containing soluble protein after expression in TBonex. C) TBonex medium kept after expression and pelleting. D) Insoluble fraction from expression without induction in YT. E) Supernatant after expression in YT. F) YT medium kept after expression and cell pelleting. All gels were stained with Instant Blue (Expedeon) dye and protein size was estimated relative to a Precision Plus Dual colour Marker.

As verified by Figure 3.15, none of the PlcH constructs (PlcH_1 - PlcH_4: two different construct lengths (82 and 73 kDa), N- and C-His tagged ((PlcH_2 and PlcH_4) and (PlcH_1 and PlcH_3) respectively) (Table 3.1) show protein secretion into the medium but all remain inside the cells they are expressed in, as no bands

are visible at the corresponding molecular weights for the media fraction. No protein of the desired size was found in media concentrated 8 times either (results omitted for clarity). Thus, expressing PlcH in *E. coli* prevents release of a potentially toxic enzyme in an uncontrolled manner and no PlcH is secreted from *E. coli* unlike the native protein, $PlcHR_2$, expressed by *P. aeruginosa*.

3.2.2.2 Large-scale expression and purification

Large-scale expression (in multiples of 500 mL) of constructs PlcH_2 and PlcH_4 in Rosetta I (DE3)pLysS cells was carried out as described in section 2.4.3.
Purification strategies outlined in 2.5 were followed (proteins were purified using an IMAC column (section 2.5.2) followed by overnight dialysis, anion exchange MonoQ 5/50 column purification (section 2.5.3) followed by gel filtration using a Superdex200 16/60 column (section 2.5.4). Protein was identified as peaks in the absorbance at 280 nm and relevant fractions were collected and analysed by SDS-PAGE.
Amount of pure, uncleaved protein usually obtained for PlcH_2 was 2 mg of protein/L culture and for PlcH_4 was 0.5 mg/L culture.



(a) (b)

Figure 3.16: 12% SDS PAGE of a peak fraction after an optimal expression and purification with gel filtration as the final step. Proteins were stained with Instant Blue (Expedeon). (a) PlcH_2; size estimated relative to Wide Range SigmaMarker^{TM} (Sigma-Aldrich) and (b) PlcH_4; size estimated relative to a Wide Range SigmaMarker^{TM} (Sigma-Aldrich). The proteins remain uncleaved.

Figure 3.16 (a) and (b) show a single peak fraction after gel purification (final purification step) from an optimal purification of uncleaved PlcH_2 and PlcH_4. Strong bands are seen at the correct molecular weights for PlcH_2 and PlcH_4 (Table 3.1) with a few fainter bands at approximately 15 kDa in Figure 3.16 (b). The results clearly indicate expression of the desired recombinant proteins albeit with some impurities.

3.2.2.3 pOPINF protein expression and purification

To rule out activity from the prepared protein samples being caused by anything other than the PlcH constructs the pOPINF vector was obtained from the OPPF. It was transformed into Rosetta II (DE3) pLysS cells and expressed and purified as PlcH_2 and PlcH_4. No bands were detectable on a 12% SDS-PAGE containing all the corresponding fractions after applying the samples to a gel filtration column. No SMase activity was detected using the Amplex® Red Sphingomyelinase Assay Kit (section 2.9). Data not shown.

3.2.3 Mass spectrometric analysis

A trypsin digest of cleaved PlcH_4 was performed as described in 2.6.1. The sample was analysed by mass spectrometry. MALDI-ToF (Matrix Assisted Laser Desorption/Ionisation Time of flight) was used in order to verify that the protein being expressed and purified was the one desired (see Appendix B, Figure B.1 for spectrum).
A Mascot search (available at http://www.matrixscience.com/) was conducted using the specific amino acid sequence of PlcH_4. Eight of the peaks seen in the spectrum were easily identifiable from the list of possible sequence fragments (m/z: 861.8; 969.3; 1040.3; 1550.3; 1675.4; 1986.4; 2082.6 and 3549.6). 23% of the PlcH sequence was covered by peptides identified in the data providing strong evidence that the protein is PlcH_4.

The protein bands on an SDS-PAGE gel of PlcH_2 and PlcH_4 were analysed by tryptic digest followed by mass spectrometry (School of Biological and Biomedical Sciences, Durham University). Analysis showed that the purified protein sample is contaminated with Hfq from *Shigella flexneri*. *S. flexneri* Hfq shares 100% identity with Hfq from *E. coli* leading to the conclusion that the contamination in the PlcH protein samples is *E. coli* Hfq.

3.2.4 CD spectroscopy on PlcH constructs

In order to verify that the expressed and purified PlcH constructs were properly folded, CD spectra were recorded and analysed as described in section 2.7.
The experimental, processed data spectra are shown in Figure 3.17.

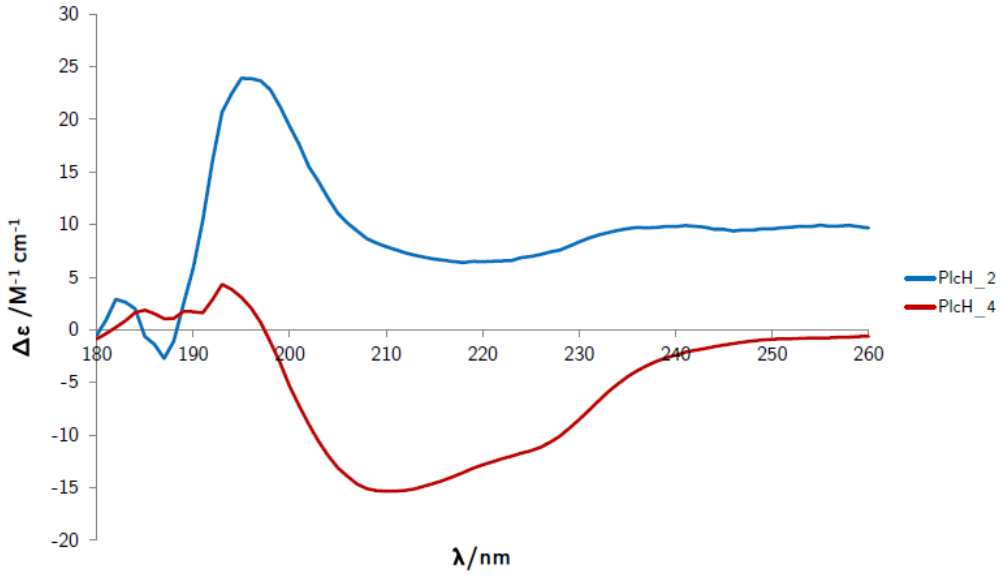


Figure 3.17: Experimental CD data, molar circular dichroism, plotted against wavelength (nm) for PlcH_2 and PlcH_4 (58 and 119 nM respectively), measured at 21 °C.

Figure 3.17 shows the experimental spectra, $\Delta\ \epsilon$, plotted against the wavelength. Qualitatively, the two spectra possess features characteristic of spectra for folded proteins with representative structures.[43,44] The spectra do not look similar, despite construct PlcH_4 making up the core of PlcH_2. The spectrum for PlcH_2 suggests that this construct mostly consists of β-sheets and the spectrum for PlcH_4 suggests that this construct may be disordered.

The data were further analysed using three different programmes - CONTINLL, CDSSTR and SELCON3 - in order to estimate the secondary structure. The analysis yielded molar circular dichroism, $\Delta\ \epsilon$, calculated by each programme (using the same basis set of proteins) which was then compared to the experimentally measured $\Delta\ \epsilon$ values (Appendix B, Figures B.2 and B.3). For both PlcH constructs, the CDSSTR programme analysis most closely matched the experimental data.
The percentages of secondary structure elements as calculated by the CDSSTR

programme is shown in Table 3.2.

Construct	α-helices (%)	β-sheets (%)	Turns (%)	Unordered (%)
PlcH_2	36	33	15	16
PlcH_4	58	18	18	6

Table 3.2: Calculated secondary structure elements for the two PlcH constructs. CDSSTR analysis showed for PlcH_2 and PlcH_4.

A more quantitative analysis performed by the CDSSTR programme calculated the distribution of secondary structure elements of the two PlcH constructs which is summarised in Table 3.2. PlcH_2 is folded evenly in α-helices, β-sheets and turns whereas PlcH_4 has a larger proportion of α-helices but fewer β-sheets. The amount of turns in the PlcH_4 construct corresponds to the first 39 residues missing compared for PlcH_2. The discrepancy between the two structures may be linked to sample quality.
The CD data shows that the expressed and purified PlcH constructs are likely to be properly folded proteins as the experimental spectra do not show the features of a denatured sample.

3.2.5 Thermal stability assay on PlcH

Various techniques exist for screening compound libraries and their effect on soluble protein; one of the simpler methods is thermal stability assays (TSAs). Techniques involving measurements of changes in fluorescence are also known as Differential Scanning Fluorimetry (DSF)[122] or Thermofluor assays (TFAs). For the following experiments a fluorescent probe, SYPRO Orange was added to the protein solution; it possesses high fluorescence when in non-polar or hydrophobic environments whereas it possesses low fluorescence in polar or hydrophilic environments.[53] The samples are slowly heated whilst the fluorescence is monitored; the change occurs as the protein unfolds as the temperature rises, leading to an exposure of the hydrophobic regions and thus an increase in fluorescence, allowing for the identification of the protein melting temperature.
The melting temperature of the protein studied can be altered, by affecting the protein stability by the addition of various additives causing changes in the protein environment. Factors usually screened for include changes in pH, salt concentrations, ionic strength and potential inhibitor compounds. This technique thus enables the identification of potential additives and buffer conditions which may enhance crys-

tallisation conditions for the protein in question.
Protein stability and solubility is dependent on its free energy of unfolding which correlates with its melting temperature. The free energy of unfolding, ΔG_U, decreases as the temperature increases and equlibrium is reached when the concentration of folded and unfolded protein are the same, resulting in $\Delta G_U = 0$. This is the parameter affected by varying the protein environment.
Plotting fluorescence measurements against temperature results in sigmoidal curves (ideal behaviour). The inflection point of this curve occurs at the temperature at which $\Delta G_U = 0$, corresponding to the melting temperature of the protein under the associated condition Figure 3.18, Figure 3.19 and Table 3.3 show the representative curves and a summary of the melting points respectively.

Conditions tested included a range of additives and salts commonly employed in protein buffers (Table 3.3) to span solutions with as large a range of ionic strengths and combinations as possible with the amount of PlcH_2 protein available. Furthermore, miltefosine was chosen as an additive and tested based on a paper by M. J. Wargo *et al*,[123] showing that this small molecule has an inhibitory effect on the sphingomyelinase (SMase) activity of $PlcHR_2$. D-*erythro* sphingosine was chosen as it is a ceramide analogue and thus resembles the $PlcHR_2$ substrate, sphingomyelin. Both were tested at various ratios relative to the protein.

Fluorescent data of a protein solution containing the additives was collected at 1 °C intervals, recording the fluorescence. The fluorescent data for the conditions resulting in most extreme changes in melting temperatures of PlcH_2 are shown in Figure 3.18 and Figure 3.19. The data for all conditions screened is summarised in Table 3.3.

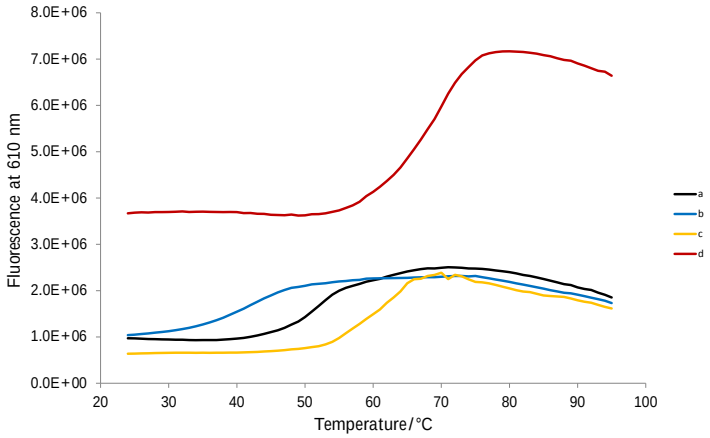


Figure 3.18: Sigmoidal curves of fluorescent TSA data. A summary of the emission of SYPRO orange mixed with $PlcH_2$ at 610 nm as a function of temperature ($24-95\,°C$). The additives are shown to be stabilising or destabilising compared to the control (no additive), PlcH_2 and SYPRO orange. Results are based on a single measurement. a) Control; b) 10 mM $NiCl_2$; c) 1 M $(NH_4)_2SO_4$; d) 10 mM $ZnCl_2$. PlcH_2 was at a constant concentration 5 $\times 10^{-6}$ M.

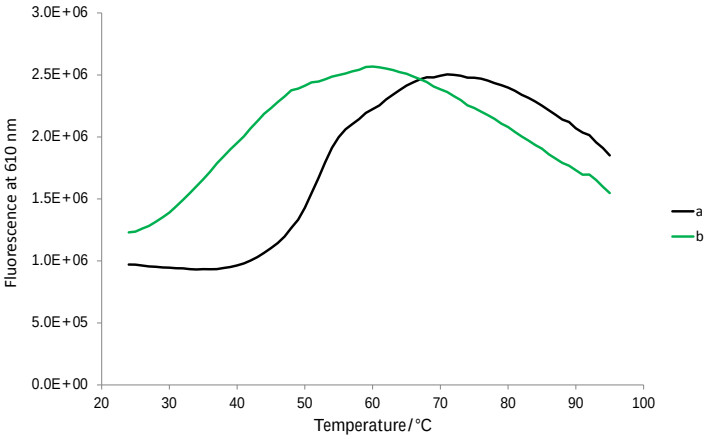


Figure 3.19: Sigmoidal curves of fluorescent TSA data. A summary of the emission of SYPRO orange mixed with PlcH_26 at 610 nm as a function of temperature ($24-95\,°C$). The additives are shown to be stabilising or destabilising compared to the control (no additive), PlcH_2 and SYPRO orange. Results are based on a single measurement. a) Control and b) 1 mM D-*erythro* sphingosine.

Compound	Concentration/M	T_{melt}/ °C	ΔT_{melt}/ °C	Ionic strength/M
Control		51	0	
NaCl	1	53	2	1
	0.4	52	1	0.4
$(NH_4)_2SO_4$	1	60	9	3
	0.4	54	3	1.2
NH_4Cl	0.4	51	0	0.4
Na_2SO_4	0.4	55	4	1.2
$MgSO_4$	0.4	50	-1	1.6
	0.01	50	-1	0.04
$CaCl_2$	0.01	50	-1	0.3
$MnCl_2$	0.01	55	4	0.3
$ZnCl_2$	0.01	66	15	0.3
$NiCl_2$	0.01	40	-11	0.3
Li_2SO_4	0.4	53	2	0.84
$NaCHO_2$	0.4	54	3	0.4
$NaCH_3CO_2$	0.4	53	2	0.4
Miltefosine	0.0001	48	-3	
	0.00001	49	-2	
	0.000005 × [protein]	50	-1	
	0.0000025	50	-1	
D-*erythro* sphingosine	0.001	38	-13	
	0.0001	43	-9	
	0.00001	51	0	
	0.000005	50	-1	
	0.0000025	51	0	

Table 3.3: Summary of the additives tested, their concentration, effect on melting temperature of PlcH_2, at a constant concentration of 5 $\times 10^{-6}$ M.

In general, most of the additives tested did not appear to have a large effect on the stability of PlcH_2, based on the single measurements performed, however a few stood out.

Amongst the salts tested, most had very little effect on thermal stability. $NaCHO_2$ (sodium formate) had a very small influence but increased melting point by 3 °C. Na_2SO_4 at 0.4 M and $MnCl_2$ at 10 mM both stabilised PlcH_2 by 4 °C each. The biggest stabilising effect was seen by $(NH_4)_2SO_4$ which stabilised the protein by 3 °C and 9 °C at 1 and 0.4 M respectively, and $ZnCl_2$ which at 10 mM increases the melting temperature by 15 °C. A large destabilising effect of 11 °C is seen fo 0.01 M of $NiCl_2$ which may be due to the fact that PlcH_2 contains a His_6-tag. The data collected for the specific binding molecules shows that miltefosine does not appear to have much impact on the thermal stability of PlcH_2. D-*erythro* sphingosine on the other hand has a destabilising effect on PlcH_2 which increases with higher concentrations of the compound; the melting temperature for PlcH_2 decreases by 9 °C at concentrations of 100 μM D-*erythro* sphingosine and by 13 °C at 1 mM. Small concentration in the range of the protein concentration appear to have very little or no effect on thermal stability.

Overall the compounds tested did not succeed in stabilising PlcH_2 significantly, other than $ZnCl_2$ and $(NH_4)_2SO_4$. Interestingly, miltefosine which inhibits native $PlcHR_2$, does not appear to have a stabilising effect on PlcH_2.

3.2.6 Crystallisation trials of PlcH constructs

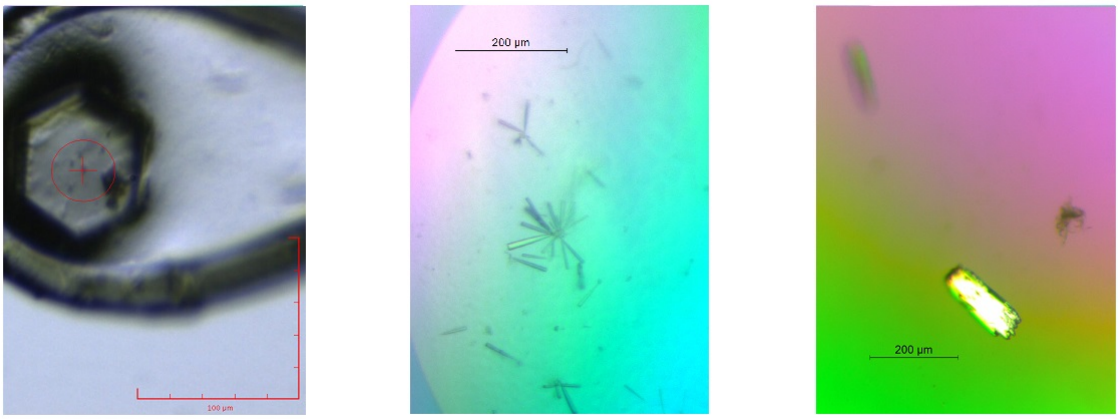
Crystallisation screens were set up for the two PlcH constructs, PlcH_2 and PlcH_4 (section 2.9.6).

3.2.6.1 PlcH_2 crystallisation

For PlcH_2, crystallisation trays were set up with the JCSG+ screen and none of the conditions tested yielded any crystals. The vast majority of crystallisation experiments showed amorphous precipitate within weeks. None of the conditions displayed any degree of crystallinity that suggested follow-up crystallisation experiments.

3.2.6.2 PlcH_4 crystallisation

For PlcH_4, crystallisation trays were set up with the JCSG+ screen.[55] Crystals were found under many conditions in the JCSG+ screen set up with PlcH_4 at 0.44 mg/mL after approximately 5 months.
Several optimisation trays were set up around conditions G9 and G11 in the JCSG+ screen, and resulted in similar crystals.



(a) (b) (c)

Figure 3.20: Pictures of crystals found in trays set up with PlcH_4 with the JCSG+ screen. (a) JCSG+ F10: 1.1 M sodium malonate, 0.1 M HEPES pH 7.0, 0.5% V/V Jeffamine ED-2001 (in loop); (b)JCSG+ G9: 0.1 M potassium thiocyanate, 30% w/V PEG 2k MME and (c) JCSG+ G11: 0.1 M Bis Tris pH 5.5, 2 M ammonium sulphate.

Several conditions in the JCSG+ screen yielded crystals of various shapes; needles (150 μm), jagged rectangles (200 μm $\times$ 50 μm), rods (200 μm) and flat hexagons (60 - 100 μm $\times$ 60 - 100 μm).

3.2.6.3 Data collection and structural analysis

Datasets were collected at Diamond Light Source (DLS) on beamline I03 for the crystals harvested from JCSG+ screen set up with PlcH_4. Crystals from various conditions diffracted but full datasets were only collected for crystals from conditions F10, G9 and G11 in the JCSG+ screen (1.1 M sodium malonate, 0.1 M HEPES pH 7.0, 0.5% V/V Jeffamine ED-2001; 0.1 M potassium thiocyanate, 30% w/V PEG 2k MME; 0.1 M Bis Tris pH 5.5, 2 M ammonium sulphate). The crystals from conditions G9 and G11 are shown in Figure 3.20.

Initial indexing of the dataset collected for the hexagonal crystals from condition F10 was undertaken using iMosflm[124] which yielded the space group and unit cell dimensions described in Table 3.4. The cell dimensions were noted as being small considering the molecular weight of 73.5 kDa of protein and a search against the known structures contained in the Protein Data Bank (PDB) showed that the dimensions were near identical to those of the Hfq protein from *E. coli*[125] (PDB code 1HK9), indicating that the diffracting protein was not PlcH_4. Molecular replacement (MR) was carried out in Phaser[126] using *E. coli* Hfq. One round of rigid body refinement was carried out yielding the R-values listed below in Table 3.4.

	Dataset
Beamline	DLS I03
Wavelength (Å)	0.9795
Lattice	Hexagonal
Space group	P_{6_1}
a (Å)	61.33
b (Å)	61.33
c (Å)	167.34
α, β, γ ($^\circ$)	90.0, 90.0, 120.0
Resolution range (Å)	52.40 - 1.86
R (%)	24.26
R_{free} (%) (based on 5% of data)	27.37

Table 3.4: Data collection and refinement statistics for the crystals harvested from JCSG+ screen set up with PlcH_4.

This established that the space groups of the collected data were identical to that of *E. coli* Hfq. No further refinement was undertaken as the R values clearly show that the structure had been solved as being identical to the Hfq protein from *E. coli* (containing 102 amino acids with a molecular weight of 11166.3 Da) and not PlcH_4.

3.2.7 Activity assay probing the SMase activity of $PlcHR_2$

The Amplex® Red Sphingomyelinase Assay Kit (Invitrogen, Molecular Probes^{TM}) was used to investigate the sphingomyelinase (SMase) activity and kinetic parameters of $PlcHR_2$ and the PlcH constructs PlcH_2 and PlcH_4 (section 2.9). Initial experiments allowed for the derivation of relevant assay parameters necessary for screening a library of ceramide analogues[127] and various commercially available compounds such as potential inhibitors of the SMase activity of $PlcHR_2$.
All experiments in this section were carried out above the critical micelle concentration (CMC) of Triton X-100.[128]

3.2.7.1 Resorufin as fluorescent standard

Assay readings were output in arbitrary fluorescent units (AFU), thus in order to express measured values in terms of enzyme activity, a calibration curve of a fluorescent dye standard of known concentration was constructed using the same instrumentation. Resorufin (Figure 3.21) is the obvious choice since it is the final product in the assay reaction. Measurements over a range of resorufin concentrations were made to construct a standard curve shown below in Figure 3.22

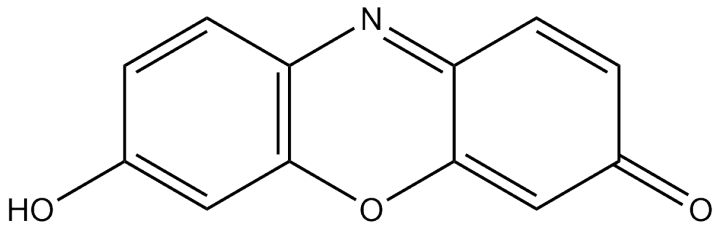


Figure 3.21: Structure of resorufin.

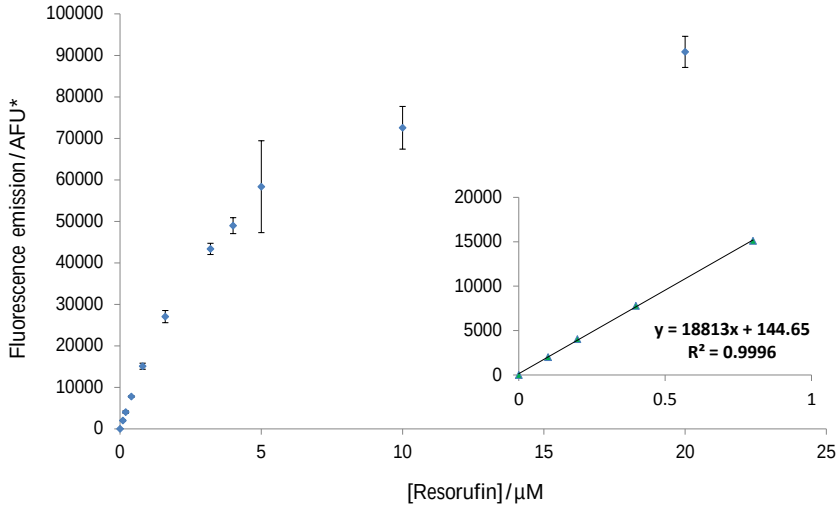


Figure 3.22: Standard fluorescent curve of resorufin at concentrations between 0 and 20 μM. The insert shows the initial, linear region. Error bars indicate standard deviations based on one set of triplicate measurements.
* Arbitrary Fluorescent Units.

A linear response was seen for resorufin concentrations between 0 and 0.8 μM as demonstrated in the insert in Figure 3.22, representing the data collected at low concentrations of resorufin. The experimental data were fitted with a linear regression curve later used to convert measured fluorescence values using the assay kit to the more commonly used unit, U.

The katal is the standardised unit (SI unit) for enzymatic catalytic activity and is defined as the enzymatic activity required for a rate of reaction with one mol of product per second under specific assay conditions (1 kat = 1 mol(product)/s).[129] The katal can not be used to express the rate of reaction, but is used to express a property of a specific enzyme. A unit (U) is used as a measure of rate of reaction; it is the amount of protein required to produce 1 μmol of product per minute of reaction (μmol min^{-1}) and so 1 U is the equivalent of 16.67 nkat. For practical reasons, one unit of enzymatic activity has been defined as the amount of enzyme required for hydrolysis of 1 picomol of sphingomyelin per minute (1 U = 1 pmol substrate/minute) under the assay conditions stated in section 2.9.

3.2.7.2 Saturation curve of the *B. cereus* SMase

In order to optimise the assay parameters *Bacillus cereus* SMase (*Bc*SMase), a standard sphingomyelinase, was used in an initial experiment. This was done to ascertain the effects of varying the *Bc*SMase concentration, between 0 and 18 nM, on the level of fluorescence detected (Figure 3.23).

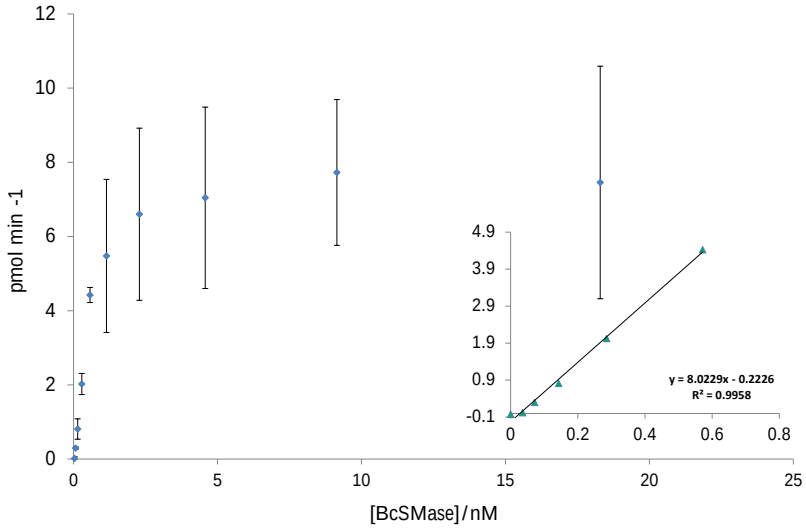


Figure 3.23: Saturation curve of the enzymatic activity of *Bc*SMase, represented by amount of SM hydrolysed, as a function of enzyme concentration. Substrate concentrations were kept constant at 0.25 mM. Error bars indicate standard deviations based on one set of triplicate measurements. Background fluorescence was corrected for by subtracting the average background (based on negative control measurements) from all measurements before averaging. Insert of initial rate region showing a linear increase.

Figure 3.23 shows a characteristic enzymatic saturation curve where the enzyme concentration is related to the rate of reaction. The curve shows an initial rate region at low concentrations where the level of enzymatic activity, and hence product formation, increase linearly with enzyme concentration. At higher enzyme concentrations, the level of specific activity rises asymptotically with increasing enzyme concentration (it approaches the theoretical maximum value).
The concentration of *Bc*SMase assayed was between 0 and 18 nM and the linear, initial rate region was determined to be between 0 and 0.57 nM (see insert in Figure 3.23). Specific activity was calculated based on the linear region of the resorufin standard curve by relating the fluorescence to resorufin concentration (the amount of

product being formed and thus the amount of substrate being hydrolysed). 0.15 nM is the molar equivalent of 1 U (1 pmol of SM hydrolysed per minute) for *Bc*SMase.

3.2.7.3 Saturation curve of the *P. aeruginosa* SMase, $PlcHR_2$

A similar experiment was conducted with $PlcHR_2$ in order to calculate the specific SMase activity.
All experiments and data processing were performed in an identical way to the one described for *Bc*SMase (section 3.2.7.2). The assay was conducted with $PlcHR_2$ concentrations ranging from 0 to 100 nM. The measured values give rise to a similar saturation curve (see Figure 3.24) as for *Bc*SMase.

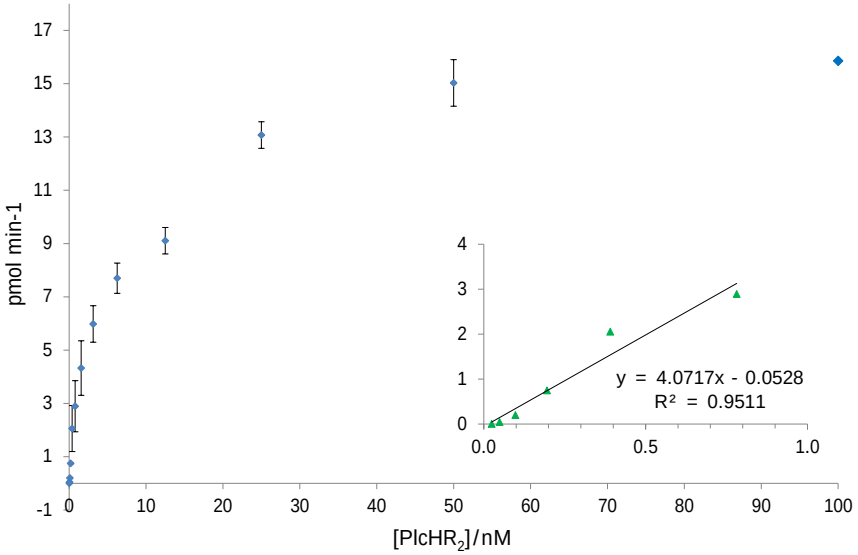


Figure 3.24: Saturation curve of the enzymatic activity of *P. aeruginosa* SMase, $PlcHR_2$. Activity measured as picomol of SM being hydrolysed per minute is plotted as a function of $PlcHR_2$ concentration. Substrate (SM) concentrations were kept constant at 0.25 mM. Background fluorescence was corrected for by subtracting the average background (based on negative control measurements) from all measurements before averaging. Insert of initial rate region showing a linear increase. Error bars indicate standard deviations based on two sets of triplicate measurements.

The linear calibration curve measured for resorufin was used to determine the enzymatic activity measured for $PlcHR_2$. A linear region was identified as the initial rate region, at $PlcHR_2$ concentrations between 0 and 0.78 nM.

This experiment establishes the relationship between protein concentration (mol/L) and enzymatic activity (U). From the initial rate region of the enzymatic saturation curve, the molar concentration of the enzyme per unit of enzymatic activity is easily deduced, employing the equation for the regression line fitted to the linear region of data (insert in Figure 3.24). 0.26 nM $PlcHR_2$ is equivalent to 1 U of enzymatic activity for the SMase, $PlcHR_2$, under the given reaction conditions.

These results were employed to determine the initial rate region, in order to ensure that all other experiment were carried out at concentrations at which the rate remains constant. The value of 0.78 nM was chosen for further experiments as it falls in the linear region where the rate of reaction remains constant.

3.2.7.4 Investigating the activity of recombinant PlcH proteins.

The PlcH constructs expressed in *E. coli*, PlcH_2 and PlcH_4, were tested under similar conditions to $PlcHR_2$ in order to assess the impact of a PlcR deletion and of the Tat-signalling sequence on SMase activity.
All experiments and data processing were performed in an identical manner to the one described for *Bc*SMase (section 3.2.7.2). The assay was conducted with PlcH_2 and PlcH_4 concentrations ranging from 0 - 100 nM for both proteins.

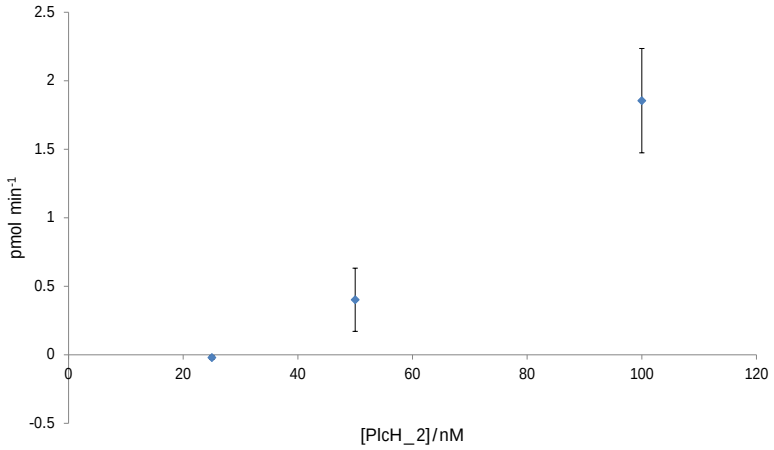


Figure 3.25: Saturation curve of the enzymatic activity of PlcH_2. Activity measured as picomol of SM being hydrolysed per minute is plotted as a function of PlcHR_2 concentration. Substrate (SM) concentrations were kept constant at 0.25 mM. Background fluorescence was corrected for by subtracting the average background (based on negative control measurements) from all measurements before averaging. Error bars indicate standard deviations based on two sets of triplicate measurements.

Figure 3.25 shows enzymatic activity of PlcH_2 as a function of its concentration. A clear SMase response is seen at high concentrations of the enzyme (at 100 nM) but at low concentrations (<25 nM) the measured fluorescent response was not discernible from the background fluorescence.
The experiment demonstrated that PlcH_2 is active as a SMase and hydrolyses SM at concentrations above 25 nM.

In experiments conducted with PlcH_4 no SMase activity was detected (data not shown). Even at enzyme concentrations of 100 nM, no activity was detectable leading to the conclusion that the loss of both the PlcR monomer and the Tat-signalling sequence as well as a C-terminal truncation result in PlcH_4 not being active as an SMase.

3.2.7.5 Enzyme kinetics of $PlcHR_2$

Many enzymes follow simple Michaelis-Menten kinetics which is based on a two-step mechanism where the enzyme first binds the substrate in a reversible reaction. The complex between enzyme and substrate is then converted into the enzyme and the product. A characteristic feature of Michaelis-Menten kinetics is that the rate of reaction approaches a maximum rate asymptotically.
In order to investigate the kinetics of $PlcHR_2$, the enzyme was kept constant at 0.78 nM whilst the sphingomyelin concentration was varied between 0 and 250 μM (0 - 13.4 mol %). The data are summarised in Figure 3.26.

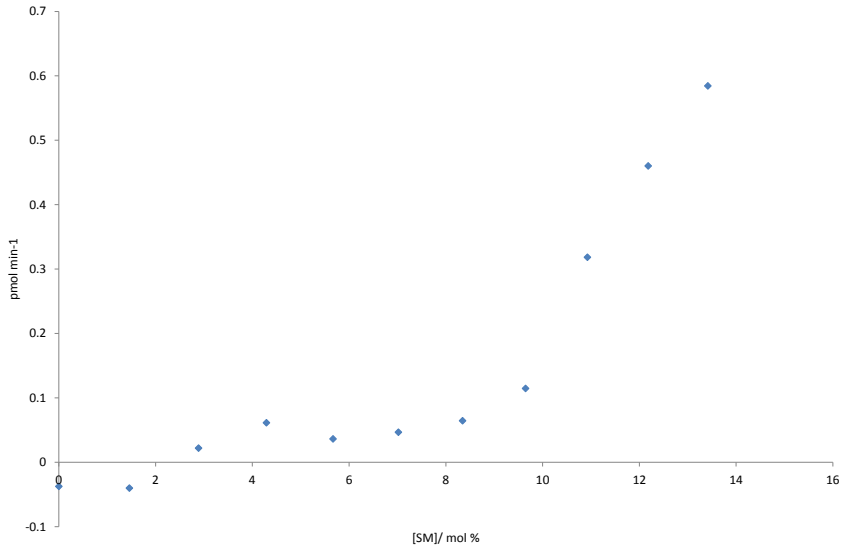


Figure 3.26: Enzymatic activity plotted against the SM mol % at a constant $PlcHR_2$ concentration of 0.78 nM. Data shown for triplicate measurement. Error bars omitted for clarity.

Furthermore, plotting the inverse rate versus the inverse substrate concentration (1/V versus 1/[SM]), a Lineweaver-Burk plot, does not result in a straight line, further underlining the fact that the enzyme does not follow Michaelis-Menten kinetics under the complex assay conditions presumably due to the fact that the substrate is presented in the form of a mixed micelle.

The Amplex® Red Sphingomyelinase assay kit is a continuous assay, resulting in the reaction carrying on until all the Amplex® Red reagent has reacted with hydrogenperoxide to form resorufin.
A timed assay was conducted at a range of $PlcHR_2$ concentrations to monitor the

substrate usage at specific times after enzyme addition in order to decipher more about the protein kinetics under the specified conditions in order to represent the reaction as progress curves until the reaction reaches an equilibrium.

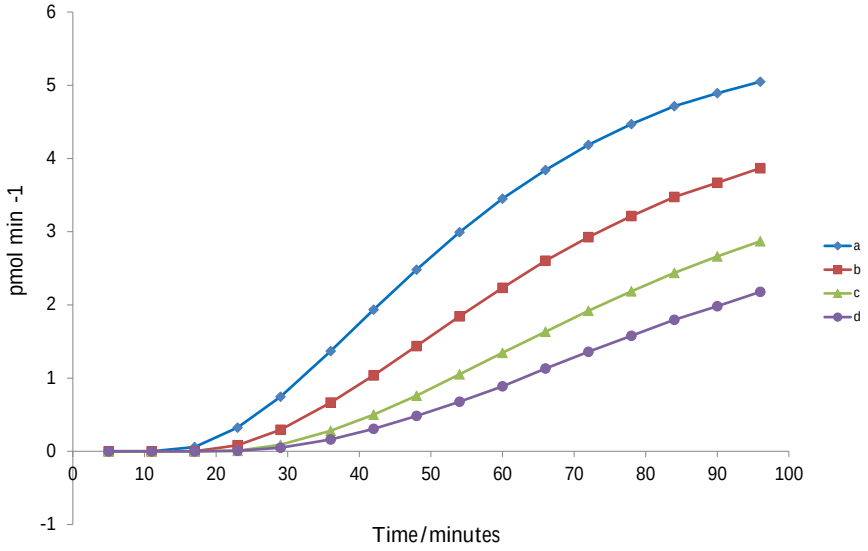


Figure 3.27: Enzymatic activity plotted against time of fluorescent measurement at enzyme concentrations of a: 2.3 nM, b: 1.6 nM, c: 0.78 nM and d: 0.29 nM. Data shown for a single triplicate measurement. Error bars omitted for clarity. Data shown for triplicate measurement.

The data from the timed assay, shown in Figure 3.27, shows a trend where the activity starts to approach a plateau at high $PlcHR_2$ concentrations as the time of the experiment progresses. This maximum rate of reaction (where the activity plateaus) is approached more rapidly the higher the protein concentration.

3.2.8 Substrate and product analogues as inhibitors of $PlcHR_2$

Having validated the assay parameters, the inhibitory effect of several substrate and product analogues was investigated as a first step towards a substrate-based structure-activity relationship (SAR).
With suitable parameters identified, the effect of these compounds on the sphingomyelinase activity of $PlcHR_2$ was investigated using the Amplex® Red Sphingomyelinase assay kit.

3.2.8.1 Ceramide analogues

Three commercially available ceramide analogues were chosen for initial inhibition assays; sphingosine ($C_{18}H_{37}NO_2$), acetylsphingosine ($C_{20}H_{39}NO_3$) and C_{17}-ceramide ($C_{35}H_{69}NO_3$), shown in Figure 3.28.

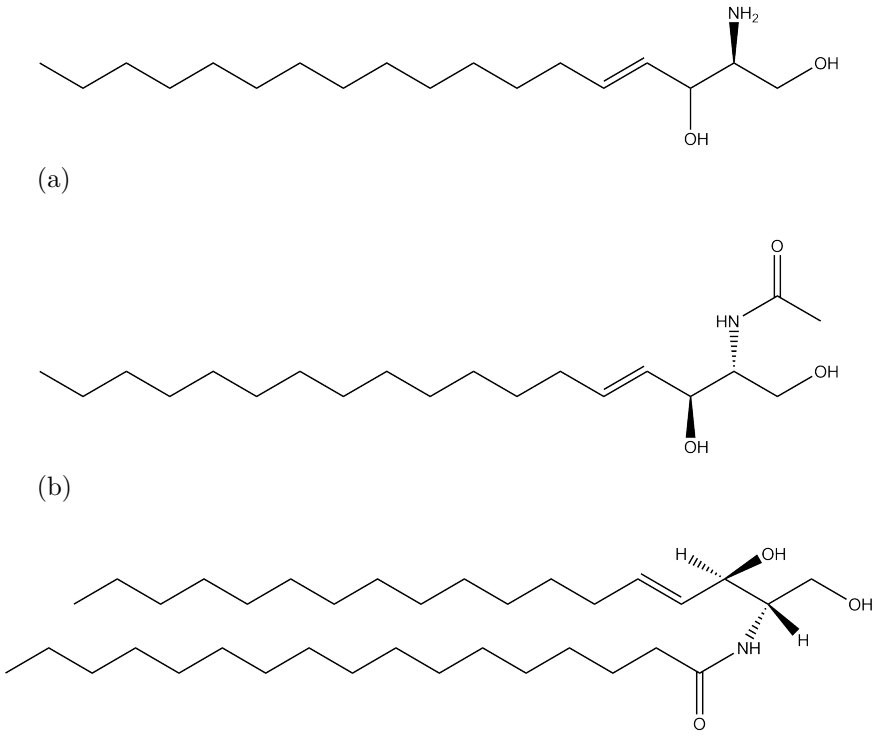


Figure 3.28: The chemical structures of the three commercially available ceramide analogues: a) sphingosine; b) acetylsphingosine and c) C_{17}-ceramide.

The $PlcHR_2$ concentration was kept constant at 0.78 nM and the SM concentration was constant at 50 μM. These parameters are based on the results in section 3.2.7.3, where the protein concentration is within the linear, initial rate region with a high enough signal (sensitivity) to ensure reproducible measurements. The inhibitors were added directly to the microplate wells to final concentrations of 0.625, 1.25, 2.5, 5 and 10 μM.

The inhibitory effect of the compounds was determined by comparing the ratio of fluorescence detected when the sample contained the potential inhibitor, protein and other reagents to a sample containing only protein and other reagents (defined as 100% activity). The ratio was then multiplied by 100 in order to express the achieved fluorescence as a percentage activity detected when specific inhibitors were

added to the reaction mixture.
The results obtained are summarised in Figure 3.29.

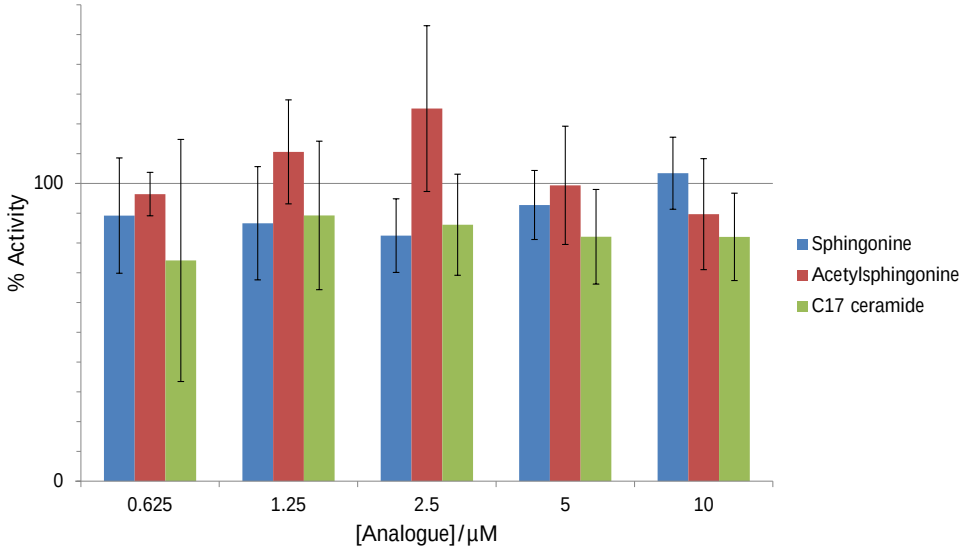


Figure 3.29: Bar chart showing the percentage activity measured for $PlcHR_2$ in the presence of 0.625, 1.25, 2.5, 5 or 10 μM of commercially available ceramide analogue (at constant protein and substrate concentrations of [$PlcHR_2$] = 0.78 nM and [SM] = 50 μM respectively). Values are averaged over two triplicate measurements. Error bars are based on standard deviations.

The results in Figure 3.29 show that none of the three commercially available analogues assayed have any significant inhibitory effect on the SMase activity of $PlcHR_2$ at concentrations between 0.625 and 10 μM. However, the data collected for C_{17}-ceramide suggests inhibition at higher compound concentration based on the small inhibitory effect, 17.9% inhibition with standard deviations of 15.9 and 14.7% recorded for compound concentrations of 5 and 10 μM respectively, suggest that higher concentrations of C_{17}-ceramide could give rise to inhibition of the SMase activity of $PlcHR_2$.

3.2.8.2 Ceramide analogue library

A library of 97 ceramide analogues (structures and numbers assigned in Appendix B, Figure B.5) was synthesised and characterised by Dr J. Mina during the course of his Ph.D.,[127] providing an opportunity to screen these compounds for potential inhibitory effect of the SMase activity of $PlcHR_2$.
Experiments were conducted as described in section 2.9. The $PlcHR_2$ concentration was kept constant at 0.78 nM and the SM concentration was constant at 50 μM. The inhibitors were added directly to the microplate wells to final concentrations of 10 μM.
The results, calculated as above, are shown graphically in Figure 3.31 and numerically in Table 3.5. The results in Figure 3.31 and in Table 3.5 only summarise the five most promising inhibitor and five most promising activator compounds. The limit was set for activities <70% and for activities >150% respectively. Only compounds meeting this requirement and with a standard deviation of <20% were included. All data are shown below in Figure 3.30 and in a 3-dimensional representation in Appendix B; Figure B.4.

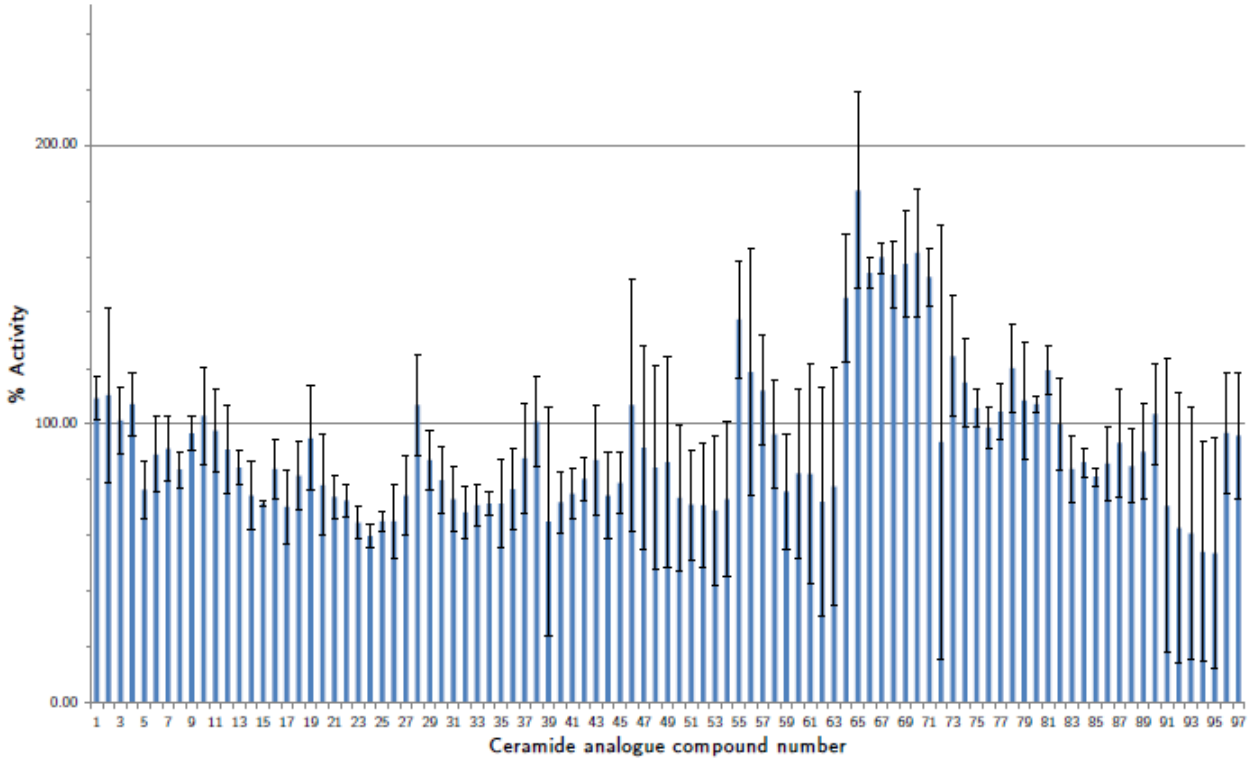


Figure 3.30: Bar chart showing the percentage activity measured for $PlcHR_2$ in the presence of 10 μM of ceramide analogue (at constant protein and substrate concentrations of $[PlcHR_2] = 0.78$ nM and $[SM] = 50\mu$M respectively) for all 97 compounds. The numbers under each bar corresponds to the the assigned compound (ref appendix). Values are averaged from 3 independent measurements. Error bars are based on standard deviations.

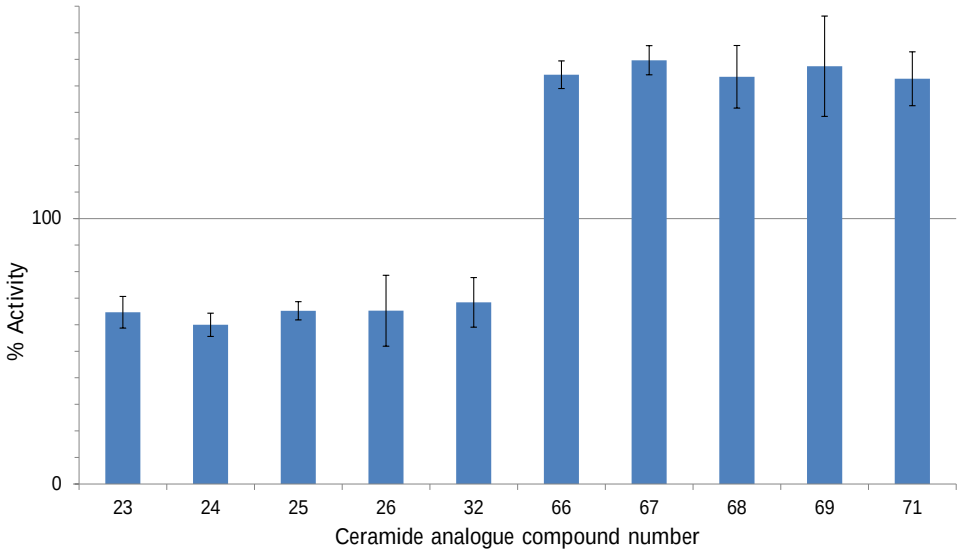


Figure 3.31: Bar chart showing the percentage activity measured for $PlcHR_2$ in the presence of 10 μM of ceramide analogue (at constant protein and substrate concentrations of $[PlcHR_2] = 0.78$ nM and $[SM] = 50$ μM respectively). The numbers under each bar corresponds to the the assigned compound (Appendix B, Figure B.5). Values are averaged from 3 independent measurements. Error bars are based on standard deviations.

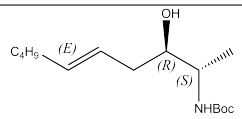
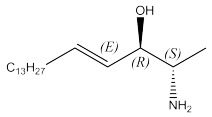
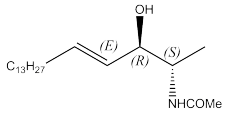
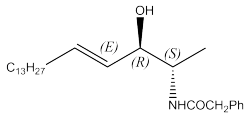
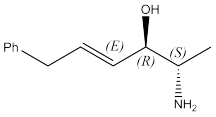
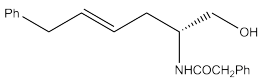
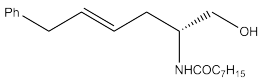
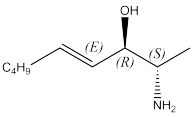
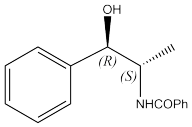
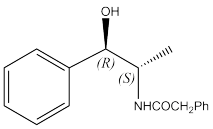
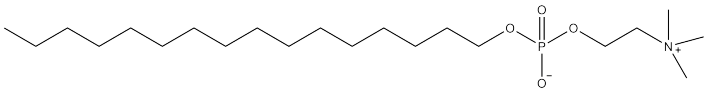
Compound number	% activity	Standard devation	Structure
23	64.7	6.0	
24	60.0	4.4	
25	65.2	3.4	
26	65.3	13.4	
32	68.4	9.3	
66	159.2	5.2	
67	159.7	5.5	
68	153.4	11.8	
69	157.4	18.9	
71	152.4	10.2	

Table 3.5: Summary of the compound number, % activity detected when compound is present, standard deviation and the chemical structure of the five strongest inhibitors and five strongest activators.

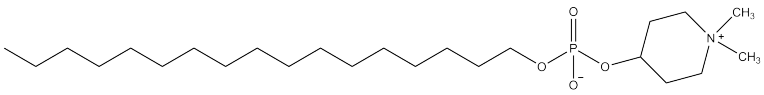
The results summarised in Figure 3.31 and in Table 3.5 show that compounds number 23, 24, 24, 26 and 32 have an inhibitory effect on the enzymatic activity of $PlcHR_2$ and they reduce the measured activity to between 60 and 68% compared to the control reactions not containing any ceramide analogue. Three of these (numbers 24, 25 and 26) contain a large, flexible carbon chain ($C_{13}H_{27}$) out of which one (24) contains a small amine group (NH_2) and the other two (25 and 26) a larger amine group (NHCOMe and $NHCOCH_2Ph$). However, the chiral centre retains the same configuration for all five compounds. Furthermore, none of these compounds contain a phenyl ring located close to the chiral centre.
Compounds number 66, 67, 68, 69 and 71 have an activating effect on the enzymatic activity of $PlcHR_2$ and increase the measured activity to between 152 to 160% compared to the control. Two of these, 69 and 71, contain a phenyl ring close to the chiral centre. Compounds 66, 67, 69 and 71 all have large substituted amines.
Out of the 97 tested compounds, five showed moderate inhibitory effects and five gave rise to significant activation of the enzymatic activity of $PlcHR_2$.

3.2.8.3 Inhibitors of sphingolipid synthesis enzymes

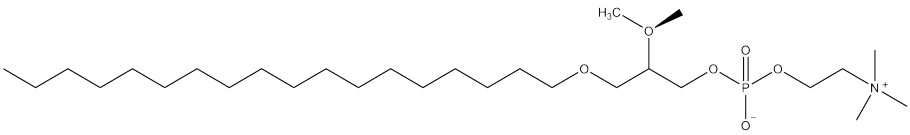
Miltefosine ($C_{21}H_{46}NO_4P$) is a commercially available compounds which possesses structural similarities to the two choline-containing lipids, PC and SM, and has been identified as an inhibitor to cleavage of these lipids[123] *in vivo*. It is currently the only orally administered drug available to treat *Leishmaniasis*. Miltefosine and two other commercially available analogues, perifosine ($C_{25}H_{52}NO_4P$) and edelfosine ($C_{27}H_{58}NO_6P$) (structures shown in Figure 3.32), were thus assayed in a similar manner to above.



(a)



(b)



(c)

Figure 3.32: The chemical structures of a) miltefosine; b) perifosine and c) edelfosine.

Experiments were conducted as described in section 2.9. The $PlcHR_2$ concentration was kept constant at 0.78 nM and the SM concentration was constant at 50 μM. The inhibitors were added directly to the microplate wells to final concentrations of 1.25, 2.5, 5, 10 and 20 μM.
The results, calculated as above, are shown graphically in Figure 3.33.

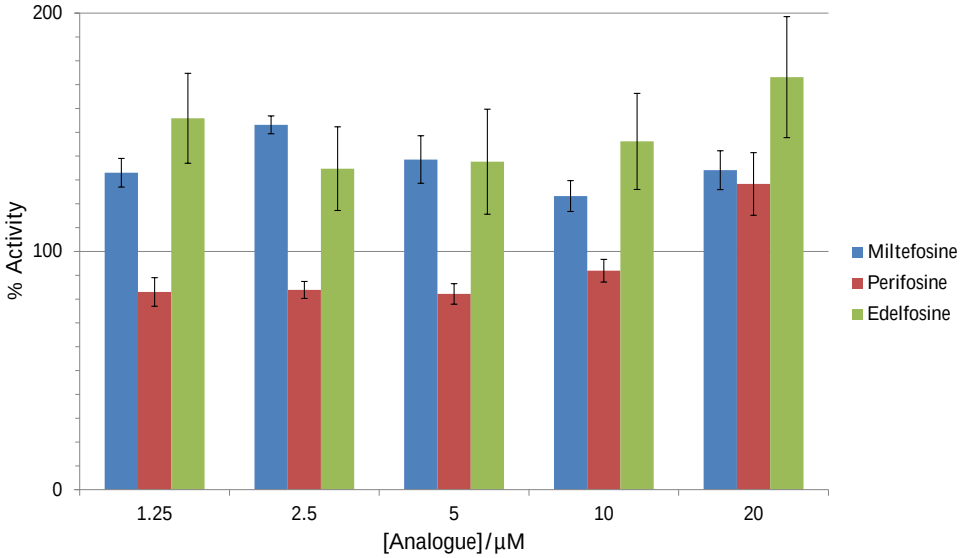


Figure 3.33: Bar chart showing the percentage activity measured for $PlcHR_2$ in the presence of 1.25, 2.5, 5, 10 or 20 μM of miltefosine (blue), perifosine (red) or edelfosine (green) (at constant protein and substrate concentrations of [$PlcHR_2$] = 0.78 nM and [SM] = 50μM respectively). Values are averaged over two triplicate measurements. Error bars are based on standard deviations.

The results in Figure 3.33 show that both miltefosine and edelfosine have no inhibitory effect on the SMase activity of $PlcHR_2$ at concentrations between 1.25 and 20 μM under the assay conditions but instead have an activating effect. The addition of miltefosine resulted in measured activities between 153% and 123% at concentration of 2.5 and 10 μM respectively and the addition of edelfosine resulted in measured activities between 134% and 173% at concentrations between 2.5 and 20 μM resepectively. In terms of structure, the two compounds are very similar; edelfosine has a longer carbon chain and contains two ether groups.
Perifosine has a small inhibitory effect on the activity of $PlcHR_2$ under the described conditions (the activity was measured to be 83%, 83%, 82% and 91% at concentra-

tions between 1.25 and 10 μM). The difference in structure between perifosine and the other two compounds is a cyclohexane ring after the phosphate headgroup containing a di-methylated nitrogen. This deviation from the typical phosphocholine headgroup may affect the inhibitory effects of this compound.
The experiments provided no evidence for inhibition of the SMase activity of $PlcHR_2$ under the conditions assayed for miltefosine and edelfosine. Perifosine showed some inhibitory effect on $PlcHR_2$ under the conditions assayed.

3.3 Discussion and future work

3.3.1 PlcH expression and purification

Small-scale expression performed at the OPPF showed that the best results expressing the highest amount of pure protein were obtained using Rosetta I (DE3)pLysS cells and the autoinduction method. All the cells tested at the OPPF, B834 and Rosetta I(DE3) pLysS, are derivatives of BL21 cells often employed for protein expression.
The first problems encountered with the PlcH constructs was attempting transformation into various cells out of which only into BL21(DE3) cells was successful. However, expression levels from those cells were below the levels of detection.
Large-scale expression was subsequently carried out with glycerol stocks of the constructs in Rosetta I (DE3) pLysS cells from the OPPF of the two most promising constructs, PlcH_2 and PlcH_4. This proved to be challenging as expression levels of PlcH_2 and PlcH_4 in both BL21(DE3) and Rosetta I (DE3) pLysS cells remained limited at approximately 1 mg/L of expression medium, which may be due to the toxic nature of the PlcH protein. Autoinduction was proved to be better for expression of PlcH than induction by IPTG addition. Utilising autoinduction media often results in a higher yield of soluble protein than when using IPTG induction due to the tight control of induction by the media composition.

An alternative to BL21 cells would be C41(DE3) or C43(DE3)[130,131] cells which are a derivative of the *E. coli* BL21(DE3) strain; plasmids that were difficult to transform into other BL21(DE3) strains due to high toxicity, have been shown to transform into C41(DE3) cells and they have proved to stabilise the plasmids during expression of toxic recombinant proteins, making these ideal for expression of recombinant PlcH. ArcticExpress(DE3)[132] cells grow at lower temperature thus slowing protein expression down, which can lead to expression of more active and soluble protein, overcoming issues with inclusion bodies.
The last obvious route to explore would be to try and express the PlcH constructs in *Pseudomonas aeruginosa* bacteria, as the levels of expression are likely to be higher in originating organism, and thus toxicity levels would not be an issue. Problems could potentially arise from purification as the constructs lack *plcR* and would therefore not be secreted from the bacteria, requiring cell lysis. It should be noted *P. aeruginosa* is a human pathogen and thus expression from this bacterium requires a biosafety level 2 laboratory which the laboratory the work was conducted in is currently not certified for.

3.3.2 Hfq contamination

Protein expression and purification was problematic due to the high levels of contamination by the *E. coli* Hfq protein found in most samples of purified protein, as verified by mass spectrometry of SDS-PAGE gel bands containing the PlcH constructs as well as crystallisation of the contaminant. The reasons for co-expression and purification of this protein with the PlcH constructs is unknown but may be due to the 24 histidine residues at the C-termini of the Hfq hexamer, allowing for co-purification on IMAC columns. Hfq is a well-known, frequently found contaminant in protein samples purified from *E. coli* and the suggested method for removal is to utilise several different purification methods.[133] However, even though three purification steps were carried out for all protein samples, the Hfq protein must have been strongly bound to the PlcH constructs, as it was not even removed after a size exclusion column. In order to minimise contamination, further purification steps could be carried out or an *hfq* gene deficient strain of *E. coli* could be used for protein expression.

3.3.3 Biophysical characterisation

In order to verify protein identity, a trypsin digest was carried out on the PlcH_4 construct. The data provided evidence that despite Hfq contamintaion, the protein was indeed PlcH_4 as 23% of the sequence was covered by the peptides identified in the data.

Analysis of the CD spectroscopy data revealed that the longest PlcH construct, PlcH_2, consists of 36% α-helices and 33% β-sheets, whereas the shorter construct, PlcH_4, consists of 58% α-helices and 18% β-sheets. These calculated values and the experimental spectra suggest that PlcH_2 is correctly folded. PlcH_4, on the other hand, does not seem to follow the same folding pattern as PlcH_2, as it appears to consist of more α-helices, despite missing the Tat-sequence as well as 43 residues at the C-terminal. The experimental spectrum also suggests that this construct may not be entirely correctly folded, but more disordered than the calculated parameters suggest.

The secondary structure elements were calculated by the three programmes; CONTINLL, CDSSTR and SELCON3, which utilise the chosen reference set to calculate values of molecular circular dichroism based on the data input into them. The data

shown for the secondary structure elements of the two PlcH constructs were calculated by CDSSTR; this was chosen by comparing the experimentally measured $\Delta\epsilon$ with the calculated $\Delta\epsilon$ by each of the programmes. The calculated data showing the best overlap with the experimentally measured values was used to predict the percentage of secondary structure elements. Even though the calculated percentages only give an approximation to the structure of the proteins, the discrepancy between the values calculated for the constructs in conjunction with the experimental spectra reveals useful information.
The protein targets, the PlcH constructs, were thus verified both by mass spectrometry and CD spectroscopy verified that the constructs were folded proteins.

3.3.4 Stabilising and destabilising conditions for PlcH constructs

Based on the results obtained when studying the thermal stability of PlcH_2 under different conditions, a few trends become apparent.
$(NH_4)_2SO_4$ and NaCl were both tested as additives at concentrations of 1 and 0.4 M; the former was found to have a stabilising effect on PlcH_2 whereas no statistically significant effect on the protein melting temperature was seen for the latter. $(NH_4)_2SO_4$ is commonly used to salt out proteins. Polyvalent anions enable salting out proteins from solution whereas polyvalent cations counteract this. Utilising a salt consisting of a univalent cation and a multivalent anion will thus increase the salting out effect. The Hofmeister series of anions is useful for predicting which salts will be effective for salting out proteins; the earlier on in this series the anion appears, the better it is at solubilising proteins. Citrate and sulphate are first in the series and thus $(NH_4)_2SO_4$ is very effective at stabilising proteins, which also explains why NaCl and NH_4Cl have very little effect on protein stability. This was further shown by Na_2SO_4 which was only tested at 0.4 M concentration but had a similar effect as $(NH_4)_2SO_4$ at 0.4 M on the protein stability whereas $MgSO_4$ had no statistically significant effect on the melting temperature.
Another important factor in protein stability is the ability of the cation to bind to the protein either via an active site or a non-biologically relevant site. $CaCl_2$, $MnCl_2$, $NiCl_2$ and $ZnCl_2$ were all used as additives at a 0.01 M concentration (ionic strength 0.3); $CaCl_2$ has no effect on stability and $MnCl_2$ only has a small effect (Table 3.3). $NiCl_2$has a destabilising effect which may be linked to the His_6-tag on PlcH_2. $ZnCl_2$ had a very stabilising effect on PlcH_2 which may be linked to its frequent use in crystallisations; Zn^{+2} may play a structural role as it is found bound both in active and non-biologically active sites in many proteins.

Miltefosine, a known inhibitor of $PlcHR_2$, was not seen to have any effect on the thermal stability of PlcH_2, whilst D-*erythro* sphingosine, a sphingomyelin derivative, destabilises the same protein significantly at high concentrations (0.1 and 1 mM) compared to the protein concentration (5 μM).
Further experiments should be undertaken including a wider range of concentrations of the already tested additives. Moreover, a buffer screen should also be carried out in order to further optimise the purification conditions of the PlcH constructs.

3.3.5 Discussion of crystallisation of PlcH

The crystallisation experiments coupled with tryptic digests of SDS-PAGE bands and subsequent mass spectrometry of those bands revealed Hfq protein from *E. coli* contamination. The contamination was not apparent in purifications and was only suspected after solving datasets. Combined with the mass spectrometry results, it has become evident that Hfq protein co-purifies with the PlcH constructs under the chosen conditions. Despite testing different purification methods and buffers, the contaminant is present in purified protein samples. The Hfq protein has been observed to aggregate into a hexamer (observed with RNA binding) and contains 24 histidine residues. Only the purest fractions after purifying on a size exclusion column were utilised for crystallisation trays suggesting that Hfq interacts with the PlcH constructs.

In order to attempt to overcome this problem, expression of PlcH constructs should be carried out in *E. coli*Δ with the *hfq* gene deleted. Another possibility is further purification on different types of columns such as a cation exchange column or a heparin column, for example.

3.3.6 Amplex® Red SMase assay

The Amplex® Red Sphingomyelinase assay kit was chosen as a suitable assay as it had been employed successfully to determine the sphingomyelinase activity of the neutral SMase from *Trypanosoma brucei*.[134] In order to establish the same assay for the heterodimeric SMase, $PlcHR_2$, from *P. aeruginosa*, various assays were carried out.
In order to gain a better understanding of the assay conditions, the theory of surface dilution kinetics is briefly discussed below.

3.3.7 Surface dilution kinetics

$PlcHR_2$ is a water-soluble enzyme which displays sphingomyelinase activity. This enables it to participate in lipid signalling pathways which include enzymes that degrade lipids (as a function of SMase activity) or participate in lipid biosynthesis (as a SMS) as well as enzymes requiring lipid activators or cofactors.[135] The environment surrounding a lipid dependent enzyme allows for different interactions to occur; three-dimensional bulk interactions and two-dimensional surface interactions. Thus, the lipid concentration in the bulk and on the surface contribute to the kinetic parameters. A kinetic model has been developed to describe the interactions at play between enzymes and lipids - 'surface dilution kinetics', by R. A. Deems *et al.*[136]

The following model was developed by R. A. Deems *et al*[136,135] and explains the reaction of a phospholipase A_2 from cobra venom ($PlLA_2$) on a mixed micelle of phosphatidylcholine and Triton X-100, a non-ionic detergent (known as a neutral dilutor as it does not have an effect on the enzyme).[137,138] Equation 3.1 represents the bulk step and Equation 3.2 represents the surface step occurring during reaction.

$$\mathrm{E{+}A} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} EA \tag{3.1}$$

$$\mathrm{EA{+}B} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} EAB \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} EA{+}\mathrm{Q} \tag{3.2}$$

where E is the soluble enzyme, which the associates with the mixed micelle, A (usually consisting of the relevant lipid substrate and Triton X-100), and forming a mixed micelle complex between the lipid, Triton X-100 and the enzyme (EA). This step is dependent on the bulk concentrations of both E and A. [A] is the sum of molar concentrations of the lipid substrate and Triton X-100, if the enzyme binds to the surface of the micelle in a non-specific way (described by the "surface binding model"). However, [A] is the molar concentration of the lipid substrate, if the enzyme during the first step binds specifically to a lipid substrate molecule in the micelle surface (described by the "phospholipid binding model"). After association with the micelle, the enzyme binds a single lipid substrate molecule, B, in its active site, thus forming another complex, EAB. This step is a function of the surface concentrations of both EA and B, as the complex is formed at the surface of the micelle. The catalysis reaction forms Q, the product, and the enzyme-micelle complex is regenerated.
The equation for the surface dilutions scheme is shown below in Equation 3.3.

$$V = \frac{V_{max}[A][B]}{K_s^A K_m^B + K_m^B[A] + [A][B]} \tag{3.3}$$

In Equation 3.3, $K_s^A = k_{-1}/k_1$, $K_m^B = (k_{-2} + k_3)/k_2$ where the former is the dissociation constant (in bulk concentration terms) and the latter is the interfacial Michaelis constant (in surface concentration units). V_{max} is the true maximum rate which occurs at an infinite bulk concentration and an infinite mole fraction of phospholipid substrate. It is assumed that the average surface area of the lipid substrate and the Triton X-100 molecules is identical; that the the surface area for each enzyme binding site on the micellar surface is constant, even if the bulk or surface substrate concentration changes; that the aggregation number and size of the mixed micelles remains constant in the concentration range utilised even when the bulk or surface substrate concentration is varied and that the enzyme has access to the entire lipid pool.[135]

Enzyme kinetics are important when attempting to understand how a specific enzyme functions.[139] Kinetics have been used to describe various enzyme reactions, mainly in solution. However, since many soluble enzymes react with aggregates of substrate (such as lipid membranes or structures consisting of proteins and nucleotides) or are membrane bound themselves, the kinetic theories are evolving to include more complex systems.

3.3.7.1 Substrate micellar structures

Several parameters are involved in the kinetics describing systems of e.g. a phospholipase and phospholipids. Firstly, the form that the substrate takes in the reaction solution is important, as rates of reactions are often dependent on the shape of aggregates the enzyme interacts with. Structures with differing morphology are formed when amphiphatic (most contain a polar headgroup and a hydrophobic tail group) molecules are dispersed in water. The aggregate structure formed is influenced by many factors including other amphiphatic molecules in the mixture, the headgroup structure and chain length, temperature, salt concentration as well as the way in which dispersion occurred. As an example, most phospholipases have been found to have preferences for specific structures and hydrolysis reactions will occur faster for certain shapes of aggregates. However, similar enzymes do not necessarily have a preference for the same type of aggregate.[139,128] Aggregates form to hide the hydrophobic tails within the structure whilst exposing the hydrophilic head groups to the water. Several studies have been carried out in order to try to

predict the aggregate shape specific substrates will form under certain conditions, employing techniques such as differential scanning calorimetry, Fourier transform infrared spectroscopy, NMR (nuclear magnetic resonance) techniques, X-ray and neutron scattering and electromagnetic microscopy (EM), making it a non-trivial issue.

Generally, amphiphatic molecules containing one hydrophilic headgroup and two hydrophobic tailgroups form structures containing bilayers such as bicelles and vesicles when in water, whereas those with only one tailgroup form micelles without a bilayer structure. If both types of compound are present in solution, both micellar and vesicle structures can form. When detergents such as Triton X-100, that usually form micelles in solution, are added to vesicles of phospholipids with long fatty acid chains, it results in the formation of mixed micelles. The detergent molecules become integrated into the vesicle lipid bilayer, which when the concentration increases eventually distorts the bilayer structure, transforming the structure to a mixed micelle. The lipid to detergent ratio necessary for forming mixed micelles is dependent on the individual lipids (differing head- and tailgroups) and detergents. The critical micelle concentration (CMC - the concentration of detergent monomer where a significant amount of micellar aggregates appear) of the detergent must be subtracted from the detergent concentration when calculating the ratio of detergent to lipid. In this case, the CMC represents the concentration where the amount of detergent molecules are available as monomers in solution and thus not at the micellar surface (0.22 - 0.24 mM for Triton X-100).[128]
In interfacial kinetics, mixed micelles are an important tool which allows the introduction of lipids into a system whilst mostly maintaining a similar environment, and they allow for easy manipulation of the surface concentration (via regulation of detergent and lipid ratios). It is important to note that the phospholipids are orientated along a surface (not a bilayer) with higher curvature compared to normal membranes.
The structure containing the phospholipids can have a strong influence on the enzyme affinity for the surface, and studies on PLA_2s has revealed mixed micelles, micelles and SUVs as the most common structures.

3.3.7.2 Interfacial catalysis

Interfacial catalysis has been defined as a reaction occurring involving at least one component which is present at an interface and is catalysed by an enzyme which can exist freely in solution or be associated with the aggregate.

There are several differences between soluble kinetics and interfacial kinetics, adding several factors into the kinetic considerations. The kinetics are influenced by the enzymes ability to attach to the aggregate surface, the substrate concentration at the surface and the environment in which the substrate is located; the two latter factors both affect the enzymatic activity.[139]

The manner in which the enzyme binds at the aggregate surface, whether it binds reversibly to the substrate molecules or any other molecules present at the surface, is irrelevant as long as the enzyme remains at the surface so that interfacial catalysis can occur. It is also possible for the enzyme to interact non-specifically with both hydrophilic interfacial regions or the hydrophobic core, or can simply be tethered ie. the enzyme is bound covalently to a moiety in or at the surface. It is important to note that the catalytic site must not interact during the binding step as the enzyme would otherwise bind to the substrate immediately and thus the kinetics would no longer be dependent on the surface concentration of the substrate. For integral membrane proteins the binding step is not included in the kinetic considerations as they are lodged within the membrane/phospholipid bilayer, and therefore do not depend on the substrate surface concentration.

In order to determine kinetic parameters for this kind of system, it is vital to understand the mechanism with which the enzyme is held at the aggregate surface. Both the enzyme and the substrate bulk concentrations must be known and whilst the bulk enzyme concentration is easily obtained, the bulk concentration of the substrate can be difficult to verify.

The enzymatic activity is a function of the bulk and the surface concentrations of the substrate. The interaction occurring between the enzyme and the surface needs to be determined to establish whether to include the bulk concentration of the substrate only or the bulk concentration of the detergent in the equation as well. The surface concentrations of the surface components (measured in mol/cm^2) define the reaction as the second step in the catalysis reaction occur in two dimensions only. This is obtained by dividing the moles of enzyme and substrate present at the interface by the total interfacial surface area, which is simplified as a mole fraction when assuming all molecules at the interface occupy a similar volume which is invariable under the conditions of the assay.

In the mixed micelle, the substrate becomes fixed at the micellar surface which affects the kinetics of the enzyme in multiple ways; the substrate can be fixed in a specific conformation which either reduces or increases interactions with the protein depending on the environmental steric interactions. This means that depending on the environment and the micellar shapes, the enzymatic activity can change drastically.

3.3.8 Sphingomyelinase activity of native PlcHR$_2$ and PlcH constructs

Native PlcHR$_2$ shows the expected enzymatic saturation curve and sphingomyelinase activity. It was shown to be slightly less active than *Bc*SMase, a standard SMase. The native PlcHR$_2$ is also approximately an order of magnitude more active than the full length PlcH_2 construct, where the latter has lost most of its activity and only shows signs of SMase activity at concentrations above 25 nM. The reasons for the loss of activity may be due to the loss of PlcR$_2$ which has been shown to contribute directly to enzymatic activity. Furthermore, PlcH_2 is folded as shown by CD spectroscopy, however some crucial parts may be flexible without PlcR$_2$, reducing enzymatic activity. PlcH_4 showed no detectable SMase activity and CD spectroscopy indicated a higher degree of unfolding.

3.3.9 PlcHR$_2$ kinetics

An assay was conducted varying the sphingomyelin substrate concentration in order to investigate native PlcHR$_2$ kinetics. As shown, the enzyme does not follow simple Michaelis-Menten kinetics under the assay conditions, as plotting the data as a Lineweaver-Burke plot does not result in a straight line. However, to establish whether an enzyme follows interfacial kinetics, employing the surface dilution model, extensive measurements must be made at various surface concentrations of the substrate and utilising various enzyme concentrations in order to determine the enzymatic activity. Due to limited materials, this experiment was only performed at a single enzyme concentration. The kinetics of the system are very complex and more data are needed in order to determine the enzyme kinetics.
A timed assay was also performed in order to study the progress curves until the reaction reached an equilibrium. This particular assay was not ideal for this study as the levels of fluorescence recorded were lower than expected after the usual 30 minutes of incubation, at the various concentrations, as the assay plate was constantly moved between 37 °C, incubation temperature, and T$_{room}$, thus slowing down reaction times. Furthermore, the platereader fluorescent detector was saturated quickly and thus unable to read the levels of fluorescence achieved at the point of equilibrium of the reaction.

3.3.10 PlcHR$_2$ ligands

When a potential inhibitor is added to to the reaction mixture, the interfacial catalysis steps are affected. The inhibitor may enhance the enzyme binding to the surface, thus displaying activating properties. To determine whether a compound is a competitive inhibitor, the surface concentration of inhibitor is used when studying the inhibition dependence.[139] When analysing the results obtained for the effect of various compounds on the SMase activity of PlcHR$_2$, some of the tested compounds stood out as being activators. According to the above description, this is likely linked to the structure of the compounds and their subsequent enzyme binding at the mixed micelle surface.

The commercially available compounds screened for inhibitory effect on the SMase activity of PlcHR$_2$ were chosen due to their structural similarity to the substrate, sphingomyelin. These compounds were shown to have little or no effect on the PlcHR$_2$ SMase activity.

The library of 97 ceramide analogues showed five compounds with an inhibitory effect and five compounds with an activating effect on the PlcHR$_2$ SMase activity. Three of the inhibiting compounds contain a long, flexible alkyl chain. Four of the activating compounds contain a phenyl ring at the end of show carbon chains and the fifth compound only contains a short carbon chain. The side groups all vary in size. This shows that there is no clear trend and not enough examples to develop a structure activity relationship.

Experiments with miltefosine, perifosine and edelfosine showed a small inhibitory effect for perifosine. However, during the course of the experiment there was an indication that miltefosine had a delayed inhibitory effect (data not shown). Due to the fact that the assay is continuous with fluorescence increasing over time, the detector was saturated when attempting to obtain a quantitative result. Pre-incubating the enzyme with the compounds did not give rise to any change in results. In order to quantify the effect, one would therefore have to utilise a system that monitors the assay over a set time period with a detector possessing a high enough dynamic range that allows detection of the high fluorescent rates achieved by the system.

3.3.11 Chapter conclusions

The work in this chapter demonstrated a biophysical and biochemical analysis of the native PlcH$_2$ and the PlcH constructs from *P. aeruginosa*, drawing on multiple techniques.

For the first time, successful cloning and small-scale expression using the autoinduc-

tion method of twelve PlcH constructs leading to expression and purification of two constructs, PlcH_2 and PlcH_4 (Table 3.1), was achieved. Expression and purification proved challenging and expression levels were between 0.5 and 2 mg per litre of expression medium. The protein folding was examined using CD spectroscopy; the spectrum for PlcH_2 showed characteristics of folded protein whereas the spectrum for PlcH_4 suggested a partially unfolded sample.

Crystallisation trials for the two PlcH constructs did not result in any diffracting PlcH crystals. However, crystallisation trays set up with PlcH_4 resulted in diffracting *E. coli* Hfq protein crystals (a common contaminant of samples expressed in *E. coli* and purified by Ni-NTA columns). This was further supported by mass spectrometry analysis of an SDS-PAGE gel band of both PlcH constructs.

The enzymatic SMase activity of $PlcHR_2$ was tested using a commercially available assay kit. Once the assay had been developed, several commercially available ceramide analogues as well as an in-house library of 97 compounds were screened for their effect on the SMase activity of $PlcHR_2$. The in-house library contained five compounds with an activating effect and five compounds with an inhibitory effect on $PlcHR_2$. The SMase activity of the two expressed and purified PlcH constructs was tested under similar conditions. PlcH_4 did not appear to possess SMase activity but PlcH_2, the full length construct, showed SMase activity, albeit considerably lower than the native $PlcHR_2$. This shows that the $PlcR_2$ chaperone is required for full activity and the full length PlcH is required for activity.

Chapter 4

The *Toxoplasma gondii* Serine Palmitoyltransferase

4.1 Introduction

Toxoplasma gondii is a protozoan zoonotic parasite. It is an important agricultural as well as human pathogen, and thus incurs a significant cost to the economy and human health. The parasite causes toxoplasmosis and can infect any warm-blooded animal including humans.[140] In fact, *T. gondii* has been suggested to be the most common parasitic infection globally.

The infection is usually symptomless in healthy individuals; however, congenital toxoplasmosis is extremely serious as it can lead to severe birth defects such as blindness and intellectual disabilities. In addition, acute toxoplasmosis affects immunocompromised patients where it often causes encephalitis or infection of various organs. It is also a large contributor to mortality rates of AIDS patients. The agricultural sector is also affected by toxoplasmosis. The food industry demands safe and high quality food and pigs are known to be a source of foodborne toxoplasmosis making farm management a priority to prevent infection.[141] Toxoplasmosis is also a large contributor to sheep miscarriages.[142] Although a number of drugs including pyrimethamine and sulfadiazine are available for treatment, the parasite can remain in the host body in its inactive form until the host's immune system is compromised for example by a primary infection. The limited efficacy of current drug treatments calls for a renewed effort to identify new potential targets.

The enzyme serine palmitoyltransferase (SPT) catalyses the condensation of L-serine and palmitoyl-CoA to form 3-ketodihydrosphingosine (KDS); the first step in the biosynthesis of sphingolipids which is essential for *T. gondii* proliferation, thus any

disruption of this pathway would slow down parasitic growth *in vivo*, making this metabolic pathway an ideal target.

4.1.1 *Toxoplasma gondii*

Toxoplasma gondii is an intracellular protozoan parasite belonging to the phylum Apicomplexa.[143] This includes many important members such as *Eimeria*, *Leishmania* and *Cryptosporidium* species, both containing enteric pathogens which can thrive in the digestive tract. Another important member is *Plasmodium falciparum*, which causes malaria and is linked to approximately 700,000 deaths (Factsheet on the World Malaria Report 2013, WHO)[144] every year. *T. gondii* is therefore the crucial model apicomplexan parasite, the advantage being that it is more easily manipulated genetically and sustains stable transfection.[143]

T. gondii is a promiscuous parasite with felidae as the definitive host.[145] *T. gondii* can infect nearly any mammalian cell as well as fish and insect cell lines *in vitro*.[141,145,146] From an evolutionary point of view it is interesting to note that infections in new world monkeys and marsupials is often acute and lethal, as they have not involved in the same environment as cats and hence have very limited immunity.[147]

T. gondii was discovered in 1908 independently by two groups of scientists (Nicole and Manceaux; and Splendore) studying the protozoan pathogen *Leishmania*.[148,149] The first case of congenital toxoplasmosis was observed in 1923[150] but not established until 1938 when an infant was born who had contracted congenital toxoplasmosis and died a month later; the parasite was identified in the infant by pathologists. Congenital infection by transmission was then discovered in many other warm blooded animals (including goats, sheep and rodents).[148] The human case of congenital toxoplasmosis was linked to ingestion of undercooked meat by the mother during pregnancy by Weinman and Chandler in 1954. Extensive infection of herbivores and vegetarians was not explained through meat ingestion and congenital transmission alone, leading to the discovery of infection through contact with cat faeces. In 1970 the coccidian lifecycle[150] and the parasitic sexual stage was found in the small intestines of the cat[151] and thus characterised biologically and morphologically.
The *T. gondii* parasite is found all over the world and has the widest range of hosts,

warm-blooded animals (mammals and birds) and humans, of any parasite.[148,150] Most hosts do not display any symptoms whereas a few contract clinical toxoplasmosis; however, the reason why certain seemingly healthy individuals develop severe symptoms is unknown.

4.1.2 The complex life cycle of *T. gondii*

The parasite life cycle is complex and consists of both sexual and asexual reproduction. *T. gondii* is a polyxenous and heteroxenous parasite which is transmitted between the various host-species via many pathways. In the 1960s the heteroxenous lifecycle was discovered. The *T. gondii* life cycle represents a number of stages including both a sexual and an asexual reproduction path (Figure 4.1). The sexual stages occur in the intestine of the primary host (carnivorous) whereas the asexual stages are mainly associated with a variety of intermediate hosts, ranging from most livestock animals to humans (omnivores and herbivores).
The three infectious stages of *T. gondii* are linked in a life cycle; the stages consist of tachyzoites (form aggregates of multiple tachyzoites known as clonal groups) and bradyzoites (found in tissue cysts), both in intermediate hosts, and the sporozoites (found in oocysts that are zygotes surrounded by a thick cyst wall) linked to a resistant stage which takes place externally to the hosts.[152]
The life cycle is summarised below in Figure 4.1 and explained in detail in the following sections, 4.1.2.1 and 4.1.2.2.

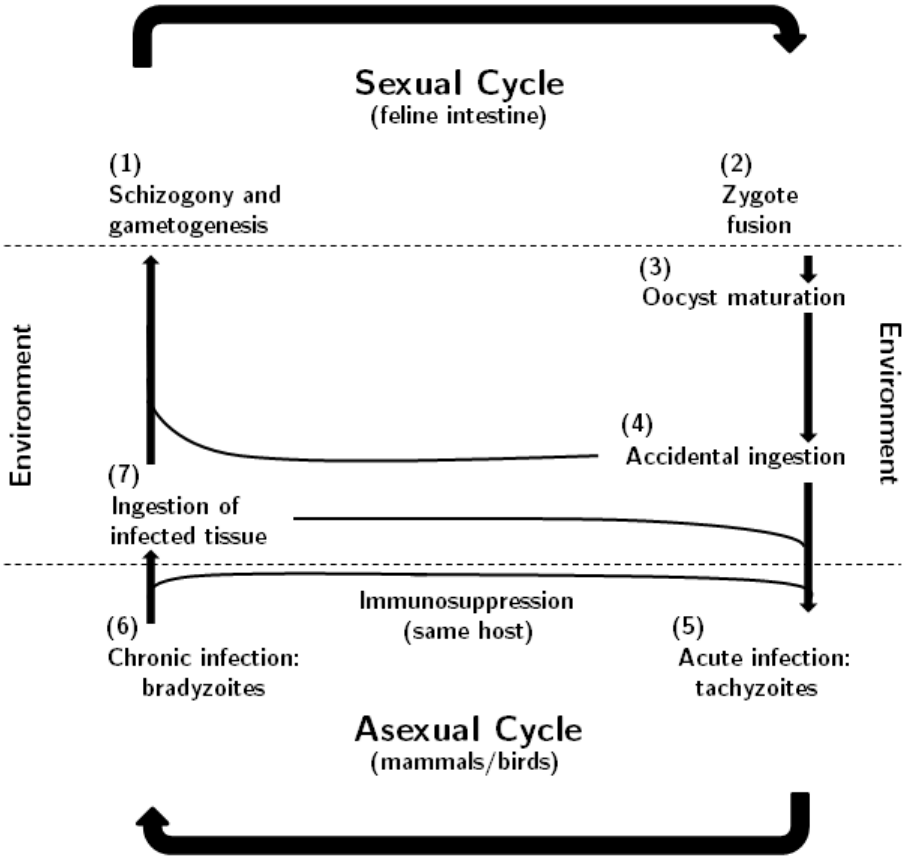


Figure 4.1: Diagram of the *T. gondii* life cycle adapted from Figure 1 in M. W. Black *et al.*[145] The diagram shows how the sexual cycle is activated in felidae only. This is followed by how the infection is spread via the environment to mammals and birds.

4.1.2.1 Asexual stages of *T. gondii*

During the acute infection in mammals and birds, *T. gondii* appears in the form of tachyzoites which are capable of rapid growth and asexual reproduction. Tachyzoites are short, crescent shaped cells with one elongated anterior and a rounded posterior (Figure 4.2), generally measuring 2 by 6 μm, containing a centrally positioned nucleus.[152] Cell movement is achieved by gliding, rotating or flexing as no physical means such as flagella or cilia are present to facilitate motion. This type of motility has been observed for certain bacteria but is more prevalent for apicomplexa.[153]

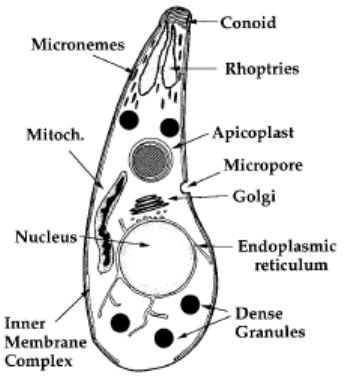


Figure 4.2: Schematic of the organelles in the *T. gondii* asexual tachyzoite stage. Reprinted from Figure 2 in M. W. Black *et al.*[145]

The tachyzoite form of the parasite can enter the host cells by phagocytosis or by penetrating through the plasmalemma where they assume an ovoid shape and become encased in a parasitophorous vacuole (PV).[154] The tachyzoite cells rapidly undergo asexual division which occurs by endodyogeny[148] within tissue cysts (PVs); the parent cell gets devoured by two progeny cells that have formed within it. This carries on until the host cell can no longer support the continued growth, at which point it bursts, thus the invasion rate is controlled by the host type. In this manner, the parasite is released around the whole host system. The parasites take this form during the acute toxoplasmosis phase[145] and tachyzoite replication only takes between 6 to 8 hours (tested *in vitro*) (Figure 4.1(5)); this is repeated until they leave the cell to cause further infection which usually involves 64 to 128 parasites aggregating in a single cell. Thereafter they differentiate and become bradyzoites (Figure 4.1(6)) which are usually found 7 to 10 days after initial infection.[145] The change can be induced *in vitro*[155] by stimuli including cyclic nucleotide kinase inhibitors or sodium nitroprusside as well as changes in pH to more alkaline conditions. Exogenous stress has been noted as an important factor for tachyzoite to bradyzoite differentiation[156,157] and may be linked to an onset of protective immunity.[156] Stress can be caused by changes in signalling pathways such as nutrient starvation or cyclic nucelotides.[155]

Bradyzoites (1.5 by 7 μm), also known as cystozoites,[148] are structurally very similar to tachyzoites; they are narrower and possess a nucleus located near the posterior end of the cell. These cells also divide by endodyogeny. They are found in tissue cysts which are most commonly found in the central nervous system (CNS) or in muscular tissues such as cardiac and skeletal muscles as well as eyes and brain, but can also emerge in other organs such as the liver, kidneys or lungs of infected species.

Cysts can contain a few or hundreds of bradyzoites and thus vary significantly in size; those found in the brain tend to be smaller than the intramuscular ones which can be up to 100 μm long - many times the size of the surrounding cells. Tissue cysts are not harmful to the host[152] and hence do not generally cause an immune response. The bradyzoites are associated with chronic parasitic infections. The parasites are believed to cross[158] the blood-brain barrier (BBB) inside leukocytes (Trojan horse mechanism) by affecting the gene expression of the endothelial cells of the brain. This has been identified for cells expressing the antigen CD11b with or without the CD11c antigen[159] and is believed to be responsible for 60% of the cells successful in crossing the BBB.[158] The asexual stage is fulfilled once the cysts are ingested (Figure 4.1(7)). Bradyzoites are released as the cysts pass through the digestive tract, where the tissue cyst walls are dissolved by proteolytic enzymes such as pepsin or trypsin,[148] but the bradyzoites released are able to survive and cause host infection. In the digestive tract the cells infect the intestinal lumen epithelia where differentiation back to tachyzoites (undergo rapid division, ready for distribution within the host) occurs, thus allowing the parasite to persist within the host indefinitely, thus continuing the cycle.

Spontaneous reactivation from bradyzoites into tachyzoites can occur; in healthy hosts, the immune system prevents distribution of the tachyzoites within the host however, immuno-compromised hosts may be unable to control this reactivation process, leading to a fatal outcome for the host.[145]

4.1.2.2 Sexual stage of *T. gondii*

Felidae are the primary host for the *T. gondii* parasite.[149] When a cat ingests a tissue cyst (Figure 4.1(7)), the cysts wall is dissolved in the stomach and the small intestine, releasing bradyzoites[152] which penetrate the small intestine epithelial cells; between the brush border and the nucleus within the PVs where they commence *T. gondii* development. The asexual developmental stages are labelled A to E (these stages are separate from generations as each *T. gondii* type consists of several generations), labelling fives stages of differing morphologies. The sexual cycle does not start until two days after tissue cyst ingestion and after 3 to 15 days, gamonts can be found in the ileum and the small intestine. The oocyst wall forms to contain the parasite whereupon the infected epithelial cells burst and the oocysts can then infect the lumen of the intestine.[149] The sexual stage involves several division stages starting two days after ingestion of a tissue cyst.[152] For 3 to 15 days after inoculation, gamonts are present in the ileum and small intestine; microgametogenesis causes

the microgamont to divide resulting in up to 21 nuclei (Figure 4.1(1)). This sexual differentiation results in the formation of microgametes which swim and fertilise the mature macrogametes by utilising their flagella, forming zygotes, whereafter the oocyst wall is formed (Figure 4.1(2)). The zygotes move through the digestive tract and are shed in faeces after rupturing from the enterocyte (Figure 4.1(3)). Sexual reproduction generally occurs by self-fertilisation.
Cats, and other felidae, can become infected through all three infectious *T. gondii* stages - tachyzoites, bradyzoites and sporozoites (found in oocysts) - which influences the rate of shedding. Oocysts and tachyzoite ingestion only lead to shedding 50% of the time whereas tissue cyst ingestion results in shedding in nearly all cats. The time between initial infection and shedding, the pre-patent period, depends on the *T. gondii* stage ingested; this period is between 3 to 10 days for tissue cysts but can take up to 18 days after ingestion of oocysts.[149] If this time period is 11 to 17 days, the ingested form is believed to be at a transitional stage somewhere between tachyzoites and bradyzoites.[148] Dubey *et al*[149] also provides evidence for bradyzoites being capable of early (after three days) development during infection (observed in cell cultures not containing any antibodies) and that conversion directly between oocysts and bradyzoites can occur.[149,152]

The only hosts excreting oocysts, the only stage that is environmentally stable, are felidae, making them the definitive host for *T. gondii*.[160] Oocysts isolated from cat faeces can be stored in normal tap water for one year and still cause toxoplasmosis in mice, in contrast to tachyzoites and bradyzoites that are killed upon contact with water.[149] They can survive storage at temperatures of 4 °C for up to a year and a half, but are killed above 60 °C, can survive at −21 °C for 28 days[160] or −10 °C for at least 106 days under laboratory conditions.[150] Oocysts have proved to be resistant to disinfectants due to their impermeable outer shell. The fact that oocysts are so resilient against environmental factors makes them almost impossible to eradicate.[150,152]

4.1.3 Infection routes of *T. gondii*

4.1.3.1 Transmission of *T. gondii* to animals

As mentioned above, cats are the definitive hosts of the *T. gondii* parasite as they are the only host that shed oocysts[149] which have the ability to survive in the environment under harsh conditions.[140] Cats can be infected via both vertical and horizontal transmission.[150] Ingestion of a single tissue cysts or bradyzoite can lead

to a cat shedding millions of sporulated oocysts in a time-period that only tends to last 1 - 2 weeks. It has been estimated that 1% of the cat population is usually shedding at one time.[140] Infection of felidae can also occur through ingestion of oocysts. Several hundred sporulated oocysts are required to infect a cat as they are fairly non-pathogenic,[150] whereas other animals such as pigs, mice and sheep can get infected by ingestion of a single oocyst.[149] In these intermediate hosts, a few oocysts can be lethal. Some species are remarkably resistant, for example, ingestion of 100 bradyzoites will not infect a mouse.
The reason contamination is so widespread in a population is due to the number of shed oocysts and their ability to survive in the environment. Domestic cats and other felidae members shed oocysts which are further spread by invertebrates such as cockroaches, earthworms and flies from the soil and even larger animals. Black bears, white-tailed deer and raccoons are all scavengers and thus a good indicator for oocyst contamination in specific areas, whereas animals used for food can be infected for years due to environmental contamination.[140]
The above examples show that nature is a ubiquitous reservoir for potential infections that can never be completely controlled or eradicated.

4.1.3.2 Transmission of *T. gondii* to humans

The most common transmission of *T. gondii* to humans occurs through contact with domestic cats and cat faeces.[140,150]
Foodborne toxoplasmosis is another route for human infection. These infections are very common and are often associated with hospitalisations and death. Toxoplasmosis has been verified as the second most frequent foodborne disease in the USA in 2012 and the fourth most frequent cause of hospitalisations linked to foodborne illness, amounting to 327 deaths and 4428 hospitalisations per annum.[160] The ingestion of *T. gondii* in food through tissue cysts in undercooked meat is a major source for *T. gondii* infection and it has has been linked to clinical toxoplasmosis. Ingestion of tissue cysts is a common transmission route to humans through ingestion of raw or undercooked meat. *T. gondii* infections are strongly affected by the cultural habits for eating raw meat and drinking unpasteurised goat milk. Thus the range of infections depend on geographical location with significantly more people infected in countries like France, where it is more common to eat raw, dried or cured meat compared to in the UK.[140,150,160] Ingestion of raw fruit and vegetables that have not been properly washed can cause *T. gondii* infection. Interestingly organic food has a higher chance of causing infection as the environment is not controlled and secured against oocyst contamination. For example, studies show that no pigs

were seropositive if kept indoors. Outdoors, free-range pigs showed that 2.9% of the ones in the study were seropositive. Infection can occur in chicken but chicken meat is often chilled or injected with enhancing solutions which mostly kills tissue cysts. Undercooked sheep-meat poses a high risk and goats-cheese from raw milk could be a contributor to toxoplasmosis. Horse meat has been the cause of severe human toxoplasmosis, especially in France.[161]

A third route of infection with *T. gondii* is via blood transfusion and organ transplant. During the latter, this can either occur through transplant from an infected donor of bone marrow or an organ into an immuno-compromised patient without immunity or be triggered in an immuno-compromised recipient patient with a latent infection. Drugs administered to lessen chances of rejection, often used in organ transplants, are both immunosuppressive and cytotoxic and can therefore trigger induction of the disease.

Generally, humans infected with oocysts as opposed to the other *T. gondii* life cycle stages display more intense symptoms and generalised disease, suggesting that oocyst may be more pathogenic.[149] Infection by tachyzoites is less important for humans,[150] however it is impossible to determine which of the other two routes is most prevalent. Due to the numerous infection routes, control of and complete eradication of infection with *T. gondii* is difficult.

4.1.4 Cell invasion by *T. gondii*

For cell invasion to occur, contact between the host cell outer membrane, containing a negative surface charge, and the parasite cell (tachyzoites; this is linked to step 5 in Figure 4.1) surface, also negatively charged, must occur.[145] Despite the parasite's ability to infect a wide range of animals, the mechanism for invasion is still not characterised. It is however believed to be due to a machinery associated with a host receptor, by insertion of the parasite receptor into the host cell membrane or by utilising one or more host cell receptors. The invasion consists of several steps starting with establishing physical contact between the host and the parasite (regulated by antigens on the parasite surface). The surface of *T. gondii* contains many proteins,[145] most of which are members of a family related to SAG1, a surface antigen, and seem to be involved in initial contact and attachment to the surface of the host cell. This step is followed by attachment occuring due to release of adhesin proteins released by certain secretory organelles called micronemes,[146] parasite gliding motility (this movement is acto-myosin dependent) on the host cell surface and penetration. The penetration is believed to be enabled by proteins known as rhoptries which are secreted by a secondary organelle; RON2, RON4, RON5 and RON8

are in a pre-formed complex which binds to the microneme protein, AMA1, to make up the moving junction complex. This junction becomes the opening to the host cell plasma membrane for the parasite. A parasitophorous vacuole is formed around the parasite where it stays whilst within the host. For tachyzoites, the microtubule cytoskeleton of the host functions by enabling cargo transport into the host cell from its periphery and movement of the rhoptry proteins which are secreted into the host.[146] It was found that by disrupting the microtubules, invasion levels are reduced if this happens soon after parasite stimulation but not if this happens at later times after invasion.

4.1.5 *T. gondii* infection

4.1.5.1 Infection in humans

Toxoplasmosis can occur in many different ways as described above. When acquired postnatally, the most common symptoms are swollen lymph nodes and occassionally flu-like symptoms (muscle aches, tiredness, sore throat, fever and headaches); this does not pose a threat to healthy humans but diagnosis is vital for pregnant women for the sake of the fetus.[140] The most dangerous infection, congenital toxoplasmosis, occurs when a pregnant women becomes infected. The most dangerous infections occur during the first trimester of the pregnancy and the effects on the fetus are more serious than if infection occurs during the second or third trimester. In most cases the infection hardly affects the pregnant woman, but the parasite within her can have devastating effect on the fetus. The starting point is a general infection, which localises in the CNS.[140] Symptoms vary across a spectrum and the manifestations vary greatly, although ocular disease frequently occurs. Mild symptoms include minor eye problems such as a decline in vision. The scale ranges to very severe problems such as retinochoroiditis, intracerebral calcification, convulsions and hydrocephalus (a build-up of fluid on the brain occurs, the least common effect).[140] Testing of all pregnant women for *T. gondii* infection is routine in some European countries, including France, Denmark, Norway, Belgium, Italy and Austria.[161] The cost-benefit of such mass screening is being debated in many other countries however, it can be beneficial to the fetus as drugs can be administered (section 4.1.5.4) to counteract potential effects.

Immuno-compromised people infected with *T. gondii* generally develop encephalitis. Often, patients undergoing treatment where immunosuppresive agents are administered, are likely to be infected with *T. gondii*. Symptoms of infection

are convulsions, drowsiness, headaches, disorientation, amongst others and many patients can become comatose. Toxoplasmosis can often be fatal to AIDS patients and even though any organ can become infected, the brain is the most commonly infected organ.[140] The majority of AIDS patients, simultaneously suffering from toxoplasmosis often experience persistent and acute headaches that can not be treated efficiently with analgesics. Headaches are usually replaced by lethargy, confusion, ataxia or even coma, during more advanced stages of toxoplasmosis. Brain lesions occur where the most predominant is necrosis.

4.1.5.2 Behavioural effect of *T. gondii* infection

Over the last decades there have been a number of reports indicating specific behavioural effects of *T. gondii* infections. For example, a recent study[162] of electronic health records found a link between cat bites and diagnosed depression in patients where 41.3% of patients admitted to hospital for treatment of cat bites also suffered from depression. Cat bites are not necessarily linked to *T. gondii* transmission but shedding is, thus cat owners are more likely to be infected with the parasite. Interestingly, the parasite has also been linked to several disorders such as obsessive compulsive disorder (OCD) and schizophrenia amongst others. Changes in personality have also been linked to infection and prenatal depression. *T. gondii* cysts are predominantly located in neural tissues, leading to the belief that it is perfectly placed to utilise its ability to thereby alter host behaviour. However, the fact that more patients with cat bites displayed signs of depression may just be an indicator of different personality types having a specific preference for pets. Whilst this is not conclusive evidence for depression being linked to cat ownership, it is interesting when considering existing studies of *T. gondii* and psychiatric disorders.

A study on infected mice showed they had higher activity levels and spent more time exploring unknown, more open areas, than the healthy mouse counterparts, leading to the conclusion that infected mice behave differently in unknown situations than uninfected ones. Studies on rats have shown that even though learning abilities may be decreased in infected individuals, it was not as pronounced as for infected mice; this was linked to the higher infection rates found in the mouse brains during latent toxoplasmosis.[163] Mice were also found to have higher mortality rates during acute infection. Due to the better health of rats, they were chosen as better models for infection. Rats are neophobic and their response to a novel situation is avoidance; a comparative study of infected and uninfected rats (both wild and laboratory rats) exposed to novel stimuli was conducted. *T. gondii* infected rats were less neophobic than uninfected rats. Further studies were made to determine the

effect of infection on the rats' recognition of the danger of cats. Rats tend to have a strong aversion to cat odour but infected rats, rather than avoiding cat-scented areas, were partial to areas with cat odour. This unnatural behaviour could, in nature, lead to higher risk for predation by cats, and thus a continued proliferation of the parasite.
This hypothesis has however been disputed by some A. R. Worth *et al*[164] who suggested that the interpretation of these findings was inaccurate. They argue that the parasite may not necessarily need to undergo the sexual cycle and that the evidence seen for enhanced predation caused by behavioural alterations is not consistent across all studies on the subject matter. Studies on the parasite *Eimeria vermiformis* have shown similar behavioural changes in infected mice, even though the parasite does not need the predatory stage for life cycle completion, suggesting that the behavioural change observed for *T. gondii* is not necessary for parasitic survival. The relative fearlessness is believed to be a side effect that has arisen due to an overall reduction of neophobia. This theory is still questioned by scientists.
In summary there is some evidence that infection with *T. gondii* can affect host behaviour, but more studies are clearly needed.

4.1.5.3 Socio-economic costs of *T. gondii* infection linked to agriculture

Various sources of infection exist on farms such as cats that shed oocysts which contaminate water and soil that the animals are in contact with as well as consumption of animal tissues, including birds and rodents, by pigs.[141,160] The incentives for avoiding animal infection are high and involve excellent farm management to keep infection pathways blocked by better cat and especially rodent control. There are also large economic costs associated with livestock infections by *T. gondii* and 1.5 - 2 million animal deaths a year could be avoided (associated with sheep and goat miscarriages).[147]
A study of migrant agricultural workers in Mexico indicated that the seroprevalence of anti-Toxoplasma IgG and IgM antibodies, was higher in workers living in rural areas compared to more urban locations. Poor workers were more likely to be infected with *T. gondii*, as their living conditions including clean water supply, sanitation conditions and healthcare services were substandard.[165] Although most healthy adults are not affected by toxoplasmosis, some display symptomatic toxoplasmosis such as dizziness, memory problems and reflex impairment, which are all contributing factors to lowered performance at work as well as related accidents.

4.1.5.4 Diagnosis, treatment and prevention of toxoplasmosis

Infection with *T. gondii* can be reliably diagnosed using various methods; the most common ones are the Sabin-Feldman dye test (a very sensitive serologic method), the direct agglutination test (DAT) (for serological diagnosis) and the detection of IgM antibodies (for detection of congenital toxoplasmosis).[148] Other examples include an indirect hemagglutination assay and an indirect fluorescent antibody (IFA) assay. However, no tests are available yet that can distinguish which infection route the host was infected through.[140]

Treatment of toxoplasmosis in humans has been treated with sulfonamides such as sulfadiazine.[140]

Pyrimethamine and spiramycin (Figure 4.3) has been utilised for pregnant women to lessen the transmission to the fetus, as these compounds are known not to cross the placenta.[148,166]

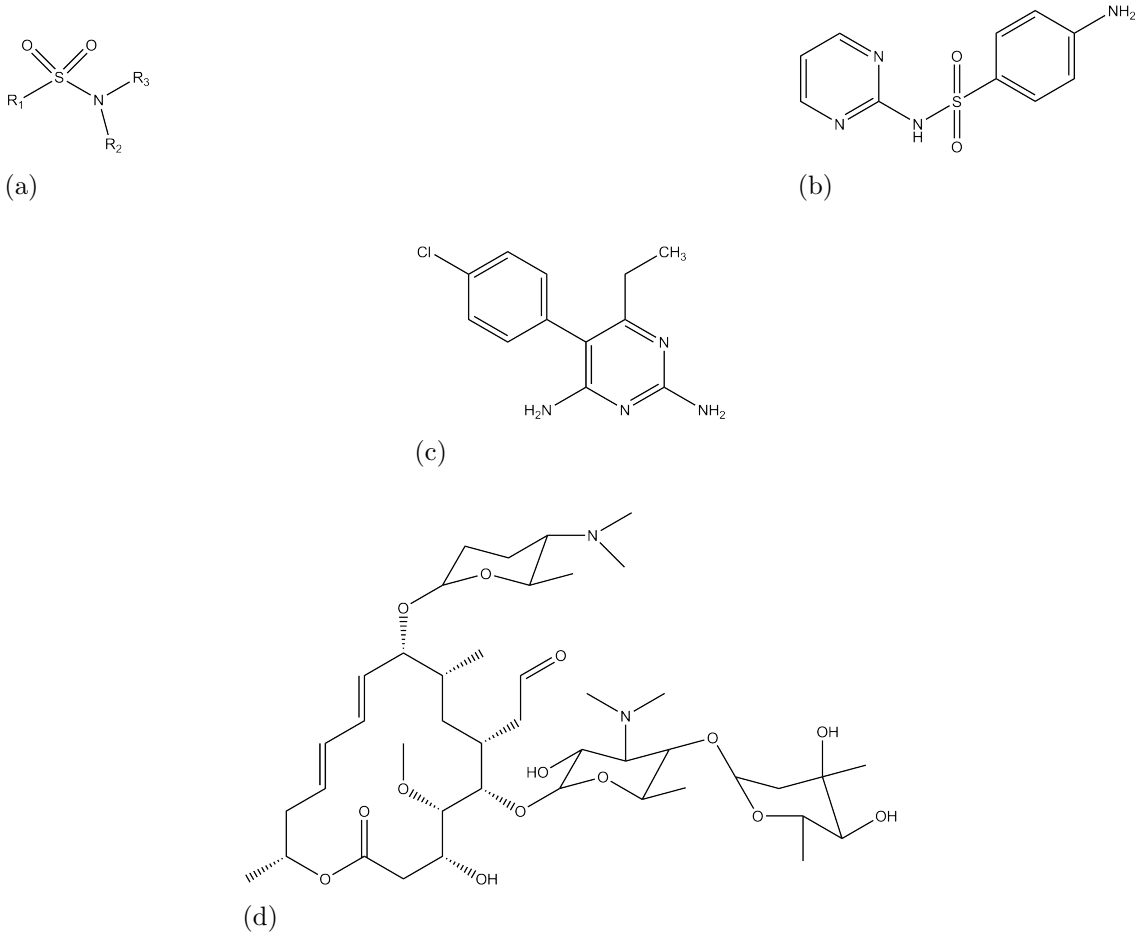


Figure 4.3: Structures of existing drugs used against toxoplasmosis. (a) Sulfonamide; (b) Sulfadiazine; (c) Pyrimethamine and d) Spiramycin.

Successful vaccinations have so far only been developed for some animals. Animals infected with *T. gondii* develop immunity against further infection, so the vaccinations currently available are live vaccines.[147] One example of a vaccine protecting against sheep and goat congenital toxoplasmosis and abortion for up to 18 months is the S48 strain Toxovax® vaccine. The strain utilised is unable to differentiate into bradyzoites which disables it from establishing a persistent infection in the host. Moreover, the sexual life cycle which usually takes place in cats, can not be initialised thereby shutting down the oocyst production route. Another oral live vaccine has been developed to prevent oocyst shedding by cats, however the downside of this vaccinations is the short shelf life and high cost associated with it.[160]

4.1.6 *T. gondii* drug targets

Apicomplexa are a type of pathogen which cause a wide range of diseases that are usually difficult to treat. They are commonly treated with chemotherapy, however new drugs targeting pathways in the parasites' metabolism such as the digestion of haemoglobin and the fatty-acid biosynthetic pathway, are emerging.[4] Often, the drug may only work against a specific parasitic life stage, complicating treatment. Another problem that presents itself is the emergence of drug resistance leading to a constant need for new drugs, treatments and drug targets.
Drug treatment of intracellular eukaryotic pathogens is challenging due to the pathogen placement coupled with the similarity between pathogen and host biology.[167] Parasitic drug targets usually consist of a specific function or component which, when inhibited, will either be lethal to the parasite or will impair its ability to function and cause infection. The target is usually a metabolic pathway, a dynamic process or a specific enzyme which does not exist within the host or differs enough to ensure that the drug will not affect the host adversely. Certain lipids are present in apicomplexa which are not found in the mammalian hosts making the apicomplexan lipid synthetic pathway a viable drug target.
The sphingolipid synthetic pathway of *T. gondii* differs sufficiently from that of the host to make it an attractive drug target.[4]

4.1.6.1 Sphingolipids in *T. gondii*

De novo sphingolipid synthesis occurs in many organisms but was reported for *T. gondii* tachyzoites by Azzouz *et al*,[168] who described *de novo* glycosphingolipid

synthesis by labelling the parasites with tritiated serine and galactose. As described in (section 1.1.2.2), *de novo* sphingolipid biosynthesis commences with the condensation between L-serine and palmitoyl-CoA to form 3-ketodihydrosphingosine (KDS), catalysed by the enzyme serine palmitoyltransferase (SPT) - a step common to all species.
More than 20 different types of sphingolipids, with both saturated and unsaturated fatty acids, can be found in *T. gondii*. It contains higher levels of sphingolipids with phosphate, with ceramide phosphoethanolamine being the most frequent one, however levels of sphingomyelin are low, all relative to the usual levels in mammalian cells. *T. gondii* also differs from mammalian cells in that the levels of free ceramides are higher, especially mono-, di- and trihexosylceramides. In order to survive and proliferate, apicomplexan parasites rely on complex lipid metabolic pathways, many of which have been identified as ideal drug targets.

Intracellular parasites can utilise a second route for sphingolipid biosynthesis, namely scavenging from the host cell to acquire lipids which are then employed to form the parasitophorous vacuole membrane usually constructed during invasion of host cells from components synthesised by the parasite or obtained from the host.[24,169,170,171]

Aureobasidin A is an antibiotic[172] which, without affecting changes to the host cell, inhibits complex sphingolipid synthesis and thus reduces the parasite's replication process, however the specific target has not been identified. 1-phenyl-2-palmitoylamino-3-morpholino-1-propanol (PPMP) and L-cycloserine (LCS) also reduce parasitic levels of ceramide synthesis.
6In plants, yeast and certain protozoa, the equivalent is inositol phosphorylceramide (IPC) produced via IPC synthase as the evolutionary equivalent.[173] This latter enzyme is not present in mammalian cells and has thus been the focus of drug targeting, both as anti-protozoals as well as anti-fungals.

4.1.7 Serine palmitoyltransferase, SPT

Serine palmitoyltransferase belongs to the PLP (pyridoxal-5-phosphate) (Figure 4.4) dependent superfamily enzymes, including the small family of α-oxoamine synthases (AOS).[50] The enzymes belonging to the AOSs catalyse the condensation reaction between acyl-CoA thioester substrates and particular amino acids, releasing CO_2 and forming an α-oxoamine.

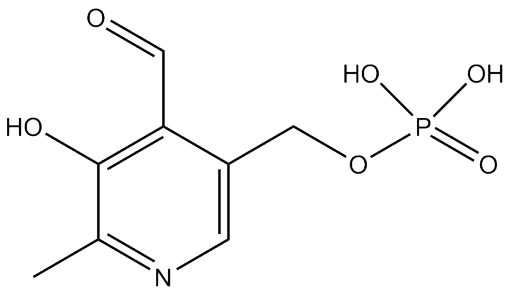


Figure 4.4: Schematic of pyridoxal-5-phosphate, PLP.

Eukaryotic SPT exists as a heterodimer consisting of long chain base 1 and 2 (LCB1 and LCB2),[174,175] which are heterologous subunit monomers. They are commonly referred to as SPT1 and SPT2 and are both vital for functional activity of the heterodimeric SPT.[50] A third SPT gene has been identified which codes for SPT3[50,176] Mammalian SPT forms a heterodimer of either SPT1-SPT2 or SPT1-SPT3. The two subunits are membrane-associated making characterisation and crystallisation a challenging task. The monomers are structurally very similar but their functionality differs significantly. Both SPT2 and SPT3 subunits contain a lysine residue in the active site which is required for binding PLP by forming a Schiff base.[176] SPT1 does not contain the conserved lysine residue as well as other important conserved residues, but instead of a catalytic role it is believed to possess a regulatory role within the SPT dimer. However, a single site mutation within a conserved region of SPT1, of a cysteine and a valine, leads to the SPT activity being down-regulated. This has been linked to hereditary sensory and autonomic neuropathy type I (HSAN1), which causes nerve abnormalities.

SPT is an important enzyme in the sphingolipid synthetic pathway as it catalyses the first and rate-limiting step for the pathway in all the organisms studied to date. 3-ketodihydrosphingosine (KDS) is sphingolipid precursor product of a decarboxylative Claisen condensation of L-serine and palmitoyl-CoA (a 16 carbon chain fatty acid). The overall reaction is shown schematically in Figure 4.5 below and described in further detail in section 4.1.8.

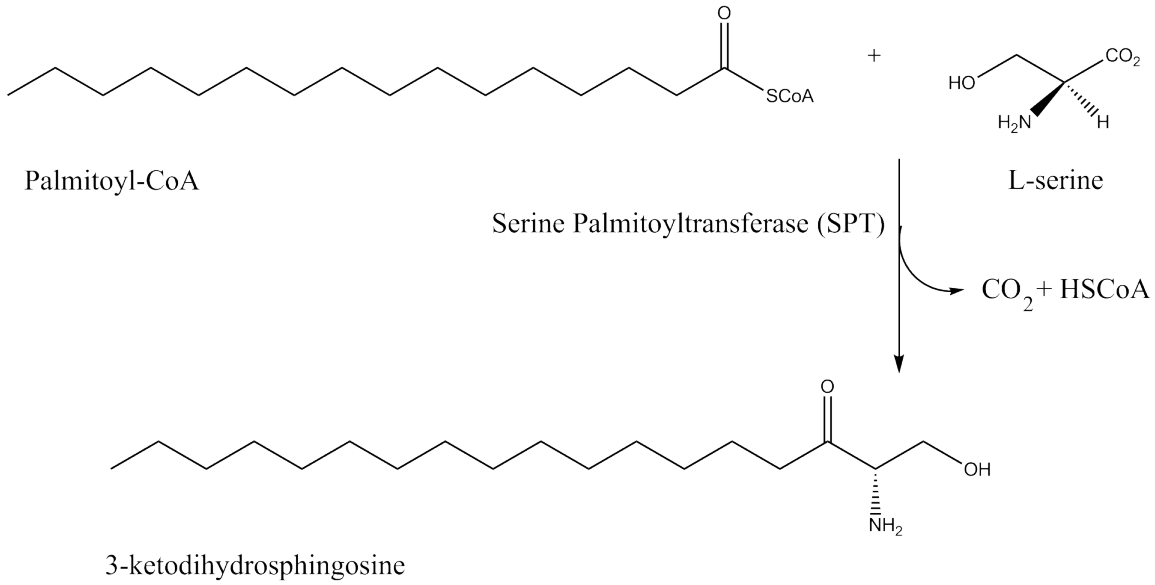


Figure 4.5: The condensation reaction catalysed by SPT.

The condensation reaction between L-serine and palmitoyl-CoA is generally used as a standard experiment to verify SPT activity (section 4.2.7).

4.1.7.1 PLP-binding site in SPT

Due to the eukaryotic SPT containing two membrane-associated monomers, isolation of the protein has been challenging, making mechanistic studies difficult. Despite the challenges posed for isolating eukaryotic SPT, small quantities have been purified from yeast and from mammalian cells. Prokaryotic SPT, on the other hand, is soluble and most mechanistic information has therefore been obtained from studies of prokaryotic SPTs. Furthermore, the active site residues are conserved from the bacterial SPTs in SPT2 and SPT3.[177] Recently, the structures of SPT from *S. paucimobilis*[50] and *S. multivorum*,[176] both homodimers, have been solved and the active site and PLP-binding residues have been identified. For *Sp*SPT the active site has been determined in the external aldimine form[50] and for *Sm*SPT, the L-serine complexed form has been determined.[176] The difference between the two structures is the active site lysine residue (Lys244), which with the L-serine forms an external aldimine, in the *Sm*SPT structure. All other active site residues overlap almost perfectly.[176] The active site asparagine residue (Asn52) is conserved in several POAS members.

The *T. gondii* SPT is evolutionarily closer to the bacterial SPTs (P. W. Denny, personal communication) and hence, it has been hypothesized to exist as a homodimer as opposed to its eukaryotic homologues.

In the *SmSPT*[176] structure the active site consists of resides from both monomers, creating a binding pocket for the substrate. The pockets are linked to the surface via long hydrophobic channels usually occupied by the long chain fatty acid part of the substrates. In the active site, the Lys244 side chain is not linked covalently to the PLP but points away from it and forms hydrogen bonds with Ser243 and Thr241 instead. The PLP pyridinium ring is stabilised by the hydrophobic interactions occurring due to the π stacking with the His138 and the Ala212. The His138 imidazole ring is positioned at the PLP *re* face. N1 in PLP is positively charged and therefore forms a hydrogen bond with the Asp210 side chain. Several hydrogen bond interactions occur between a PLP oxygen atom and His213 which then interacts with the hydroxyl group of Tyr54. The phosphate head group of PLP is stabilised by polar interactions with the residues of the active site. These include the Phe114 and the Ala274* N atoms as well as the Ser243 and the Ser273* side chains (where * denotes residues located on the other monomer). The PLP phosphate group forms hydrogen bonds with the L-serine hydroxyl group which forms an interaction with Ser243 via a water molecule. It also interacts indirectly with other residues through water molecules and it is tilted 80 ° against the pyridinium ring of PLP due to His138 interactions. This is shown in Figure 4.6.

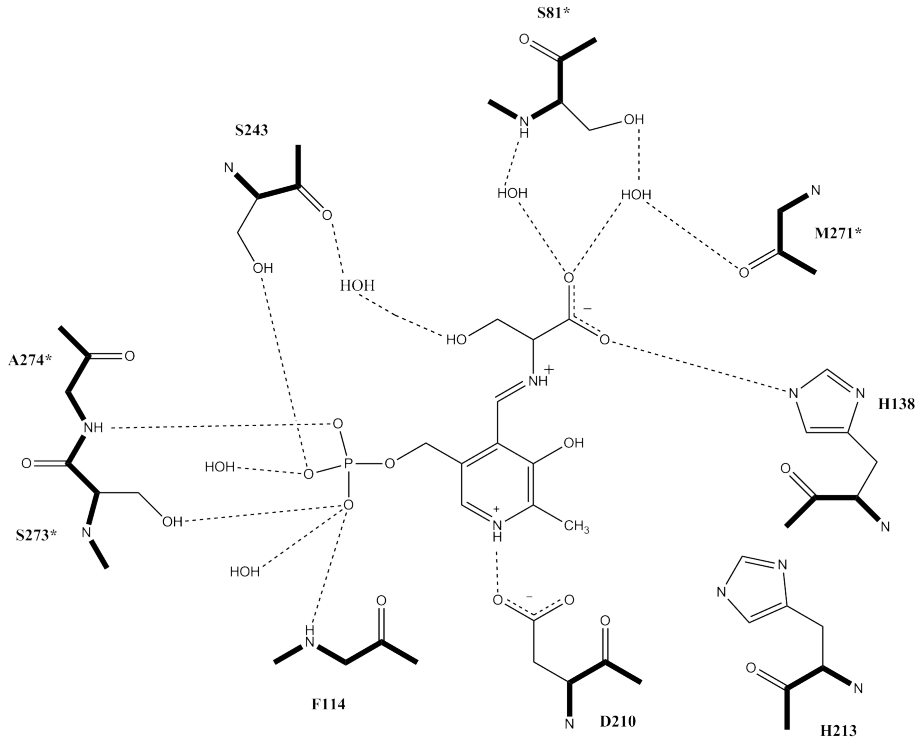


Figure 4.6: Schematic of the active site in *SmSPT* showing the hydrogen bond network, the external aldimine between PLP and L-serine and the active site residues. Water molecules are included. Residues denoted with a * are situated on the second monomer. Schematic made in Chemdraw Std. 13.0 based on Figure 5 C from H. Ikushiro *et al.*[176]

In the crystal structure of *Sp*SPT[50] PLP is bound at the active site. It forms an internal aldimine, also known as a Schiff base, with the active site lysine (Lys265). The side chains of Asn138, Asp231, His234 and Thr262 as well as the Gly134 and the Tyr135 main chains form polar contacts with the cofactor. The His159 imidazole ring stacks with the PLP pyridinium ring. The histidine is part of a three residue conserved motif, in this case His159, Ala160 and Ser161, which is conserved in the AOS family.

It is clear from Figure 4.6 that the PLP cofactor is bound by several residues in the active site which are vital for the catalysis reaction to occur. Sequence homology studies of SPTs from various organisms clearly show that these residues are highly conserved in the SPTs,[178] even though the binding sites may differ depending on

the surrounding enzyme.

4.1.8 The mechanism of action of SPT

The reaction mechanism of SPT is shown in a simplified schematic representation in Figure 4.7.

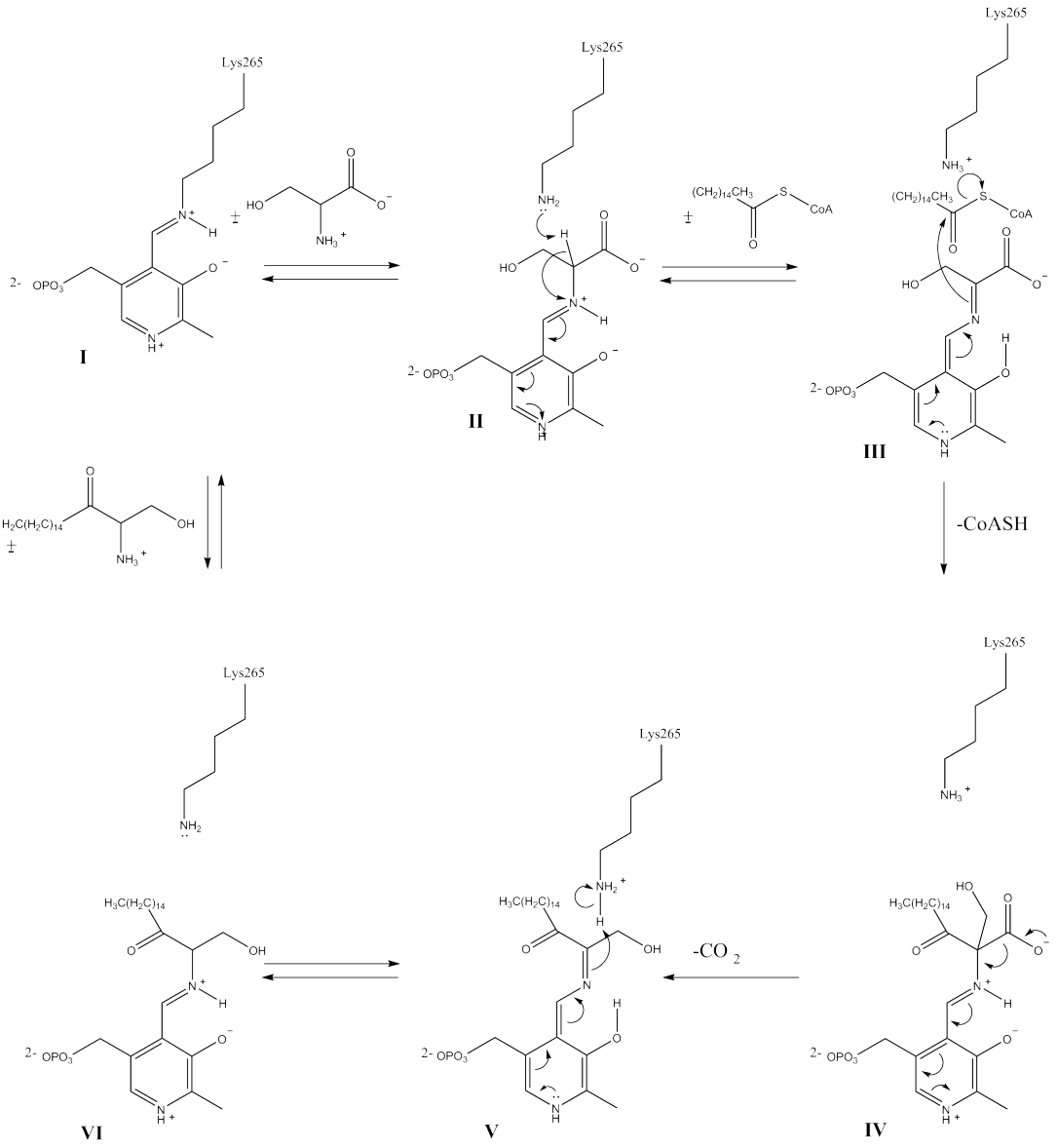


Figure 4.7: The SPT reaction mechanism. Drawn based on scheme 1 from Y. Shiraiwa *et al.*[179]

The PLP forms an internal aldimine at the active site lysine (Figure 4.7, I) which undergoes a transaldimination reaction upon addition of L-serine. This results in the PLP-L-serine aldimine, the external aldimine (II) which releases PLP

from the lysine. The external aldimine is held in the active site by polar interactions and salt bridges. The *re* face of the PLP molecule is stabilised by the conserved histidine residue which also interacts with the serine carboxyl group, which in turn prevents unwanted side reactions from occurring. This is due to the Cα-H bond with the serine being forced into an almost planar orientation in relation to the PLP pyridinium ring, which does not favour dissociation. Palmitoyl-CoA enters the hydrophobic pocket where the histidine interactions with the carboxylate are interrupted. A hydrogen bond is formed between the histidine and the acyl carbonyl O (the thioester O).[180] The L-serine carboxyl group interacts with the arginine residue which allows for the Cα-H bond to move into a more favourable, perpendicular position relative to the PLP pyridinium ring. This allows for α-deprotonation resulting in the formation of a quinoid intermediate (III). A Claisen-type condensation occurs as the palmitoyl-CoA is attacked by the Cα carbanion where the histidine residue and the fatty thioester act as an acid catalyst. This generates an intermediate (IV) whereupon the acyl-serine conjugate is decarboxylated to form a second quinoid intermediate (V). The second quinoid Cα is protonated resulting in the PLP-KDS external aldimine (VI), which releases KDS to regenerate the original internal aldimine (I).

4.1.8.1 SPT inhibition

A number of SPT inhibitors have been identified.[181] The two cycloserine enantiomers are irreversible inhibitors of a lot of PLP-dependent transaminases,[182] as they are cyclic serine/alanine analogues. D-cycloserine (DCS) is a naturally occurring compound (found in *Streptomyces* bacteria) whereas L-cycloserine (LCS) is a synthetic compound (both shown in Figure 4.8). The former is often employed in combination therapy for tuberculosis infections but only as a second-line drug as it has severe side effects. Cyclo-serine acts in an irreversible manner due to its ability to form an adduct with PLP in the the various AOS enzymes which utilise PLP as cofactor. Experiments by J. Lowther *et al*[182] have shown that 79% and 83% of the original activity of DCS and LCS treated enzyme respectively, can be restored when dialysing inhibited SPT against a buffer containing 25 μM PLP, clearly showing that these compounds bind to the PLP cofactor but do not otherwise affect the enzymes. The compounds undergo an aromatisation mechanism whereby the PLP cofactor in a holo-enzyme forms an external aldimine with the cycloserine. The aldimine is deprotonated resulting in 3-hydroxyisoxazole-PMP adduct which is stable. This reaction occurs faster with DCS than LCS, which may be the reason why the former functions better as an antibiotic.

The most commonly employed inhibitor of SPT activity, however, is myriocin (also known as ISP-1 and thermozymocidin ((2S,3R,4R,6E)-2-amino-3,4-dihydroxy-2-(hydroxymethyl 14-oxo-6-eicosenoic acid) (Figure 4.8).[183]

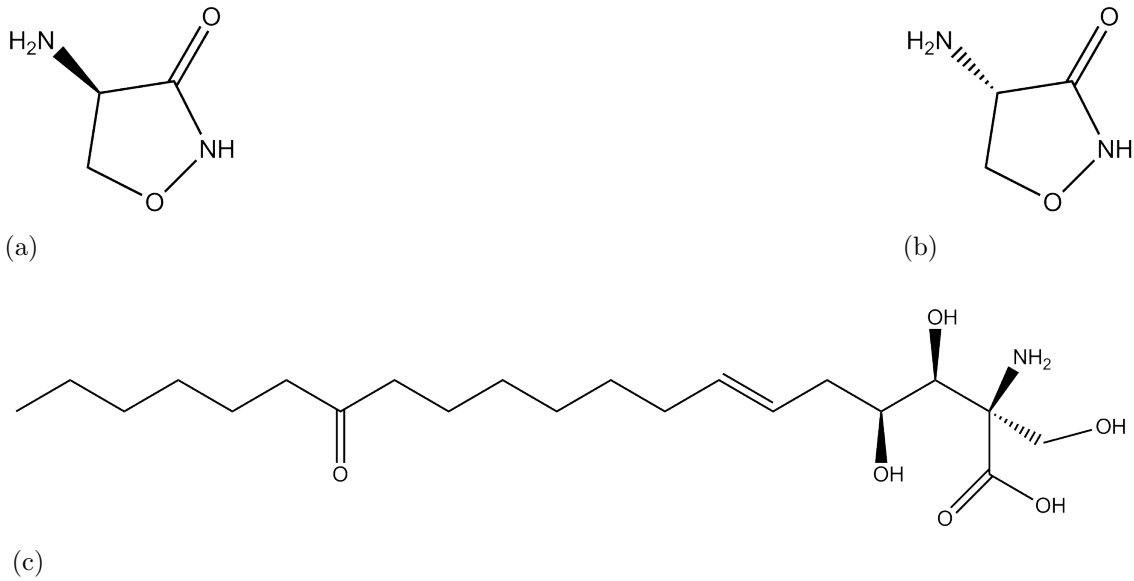


Figure 4.8: Schematic of (a) D-cycloserine; (b) L-cycloserine and (c) myriocin.

It is a natural product which has been isolated from *Mycelia sterilia* and *Myriococcum albomyces*;[184] it has been used as an antibiotic and as an antifungal agent. Myriocin is a sphingosine analogue and its inhibition of SPT has been shown to be non-competitive.[183] It is the only commercially available SPT inhibitor resembling sphingosine.[185] Inhibition of SPT with myriocin is generally thought to be non-competitive however L-serine binding to *Sm*SPT can be competitively inhibited with myriocin. Myriocin has a higher affinity than LCS and binds tightly as the chain is believed to interact with the site for palmitoyl-CoA binding. Myriocin is analogous to KDS with the differences of keto groups, the reduction of the β-keto group, a hydroxyl and an unsaturated bond which is situated at the aliphatic side chain.[185]

*Sp*SPT was employed as the model enzyme to elucidate the mechanism of inhibition of SPT by myriocin. J. M. Wadsworth *et al*[184] suggest that myriocin acts as a compound possessing dual mode inhibition of SPT.
Myriocin binds by forming an external aldimine with SPT containing a PLP cofactor (holo-SPT) (confirmed by mass spectrometry and spectroscopy experiments). Full inhibition of SPT was not recorded and is believed to be due to myriocin being very insoluble and so high enough concentrations to saturate the binding site can not be reached (although this is believed to be linked to the *in vitro* nature of the

experiments). PLP usually forms a Schiff base external aldimine with the myriocin molecule and normal SPT activity can be re-established with addition of PLP. However, if myriocin is incubated with the holo-SPT enzyme for a longer time-period, this is no longer reversible. Experiments have identified the conserved active site lysine as being crucial to the breakdown mechanism which can occur where the external aldimine is transformed into a C18 imine adduct of this specific residue (via a retro-aldol degradation resulting in a C18 aldehyde), causing the inhibition to become irreversible.
Crystal structure determination has revealed that myriocin is decarboxylated resulting in a PLP-decarboxymyriocin aldimine.

4.1.9 Chapter objectives

The project objectives were to determine the smallest catalytically active domain of SPT1 from *T. gondii*, achieved via different constructs. 48 constructs of various sizes and in different expression systems of *Tg*SPT were designed in order to optimise expression levels of the protein whilst also spanning a wide size range in order to identify the catalytic ability of the enzyme. The focus of the project was to express, purify and characterise the most promising constructs with the ultimate aim being crystallisation of one or more of them. After optimising expression and purification, the nature of the enzymes was determined using mass spectrometry and protein folding was investigated using CD spectroscopy. Bio-SAXS was employed to estimate the approximate size of the various constructs. Enzymatic activity was determined using a well-established assay utilising [C^{14}]-labelled serine and palmitoyl-CoA. Finally, various conditions of pH, buffer systems and salts were tested with using a thermal stability assay followed by crystallisation trials.

4.2 Results - Identification and characterisation of the putative SPT from *T. gondii*

This section describes the work towards a crystal structure determination of SPT from *T. gondii*; from cloning and expression to a detailed biochemical and biophysical characterisation of the protein.

4.2.1 Bioinformatical analysis of the predicted *spt* gene

In order to identify the conserved catalytic domain of the *T. gondii* SPT, the full-length sequence was analysed using a range of available bioinformatical tools. A sequence comparison using BLAST[186] (Basic Local Alignment Search Tool) revealed that the protein consists of three distinct domains. These are a transmembrane helical domain, a pyridoxal-5'-phosphate binding pocket (an aminotransferase domain) and a functional domain containing the active site.
The 47 amino acids at the N-terminus contain a transmembrane helix as identified with TMbase.[187] This helix is presumably responsible for anchoring the enzyme in the membrane. Residues 48 - 181 are not well characterised but have been identified as mainly helical using secondary structure prediction algorithms such as JPred3.[188] This part of the sequence shows no significant identity with any other functionally characterised sequence. The remaining residues 181 - 571 show significant sequence identity with a wide range of PLP dependent amino transferases.
Based on our sequence analysis shorter constructs were chosen to improve expression and prioritised in order to identify the shortest functionally active domain. The different starting points for the four constructs discussed in this chapter (Table 4.1) were based on the secondary structure predictions which indicated a number of helices. Each starting point was hence chosen to be in the random coil regions between predicted helix regions.

4.2.2 Homology modelling of the *Tg*SPT structure

*Tg*SPT shares between 20 and 30% sequence identity with a number of bacterial orthologues, some of which have been structurally characterised by X-ray crystallography.[50,176] *Tg*SPT and *Sm*SPT (pdb entry 3A2B),[176] the closest orthologue, share a 27.8% (score given in ClustalW)[189] identity. A sequence alignment of the two homologues was constructed in ClustalW.[189] The sequence alignment was utilised to generate a homology model of the 3D structure of *Tg*SPT based on

the already known one of *Sm*SPT[176] using the MODELLER software[190] (available on http://salilab.org/modeller/). The *Sm*SPT monomer was used as a template to model the *Tg*SPT as a monomer without the N-terminal residues (*Sm*SPT consists of 399 amino acids compared to *Tg*SPT consisting of 571 amino acids).
A dimer was generated in Coot[191] based on the known crystal structure and a PLP molecule was added to each monomer at the modelled active site.
PLP was built into the structural model in Coot[191] where distances to the surrounding residues were characterised in order to verify that the predicted residues (determined through sequence alignment) were in the correct conformation for hydrogen bonding to occur with the PLP molecule. The full sequence of both SPTs is shown aligned with the secondary structural elements of *Sm*SPT overlaid in Figure 4.9. It is clear that the N-terminal domain of *Tg*SPT can not be modelled as conserved residues between the two sequences are further toward the C-terminal end of *Tg*SPT. The functionally active domain and the C-terminal domain show how well conserved the residues are.

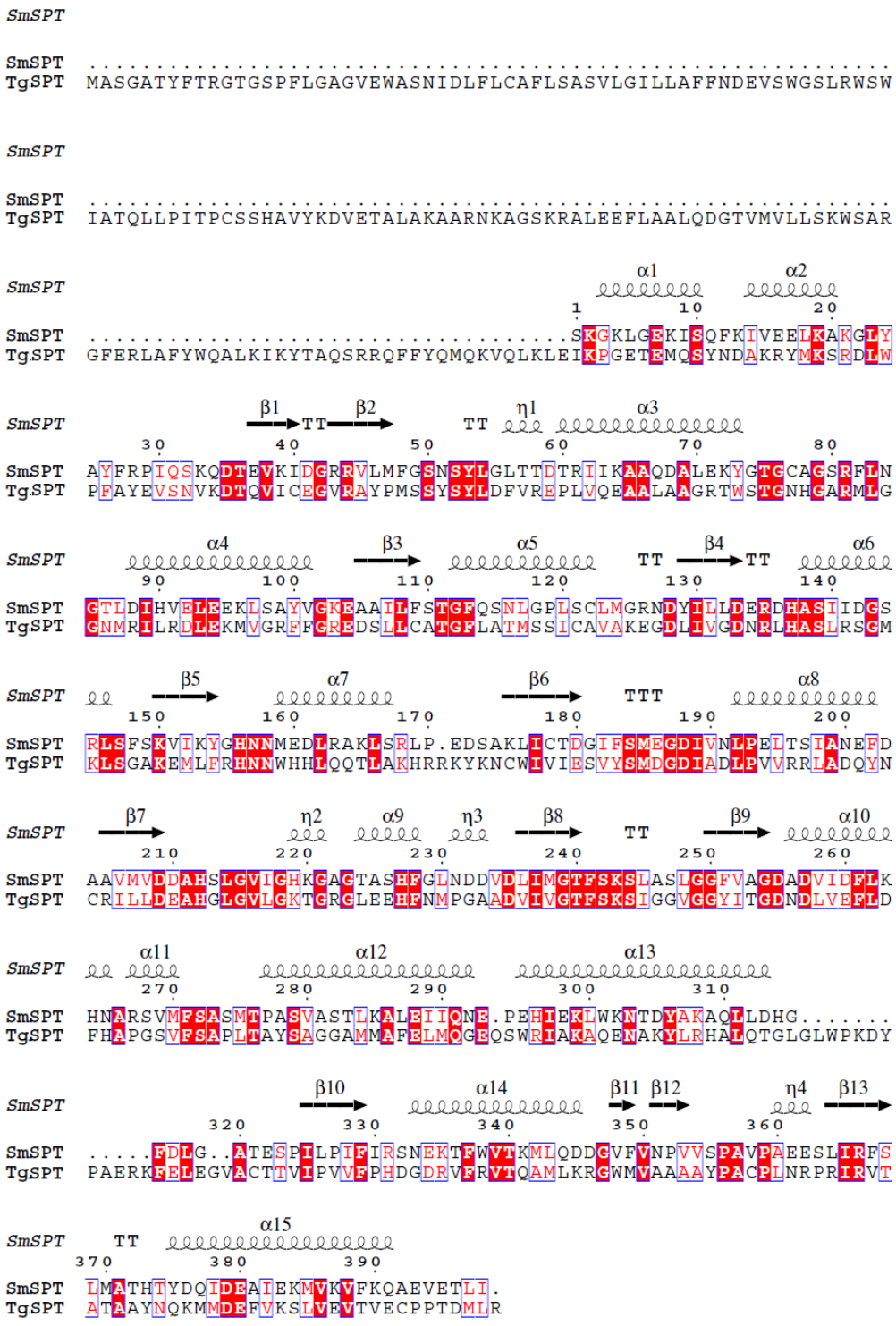


Figure 4.9: Sequence alignment of the full *Tg*SPT and *Sm*SPT sequences made in ClustalW[189] and ESPript 3[121] (available at http://espript.ibcp.fr) with overlying secondary structural elements based on the crystal structure of *Sm*SPT.[176] Residues enclosed in a red block are identical, residues enclosed in blue frames with red letters are of similar type and the rest have a low identity score. α denotes α-helices, β denotes β-sheets, TT denotes strict β-turns and η denotes 3_{10}-helices.

The 3D structural model was generated for the conserved catalytic domain of *Tg*SPT (residues 156 to 563). In Figure 4.10, the structure of the homology model monomer containing a PLP molecule in the binding site is shown. The fold is mainly α-helical with a binding cleft for substrate binding.

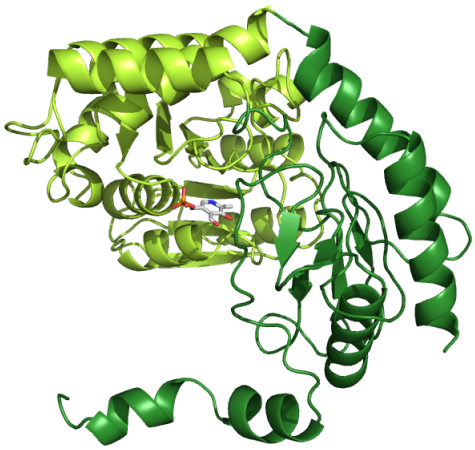


Figure 4.10: Overall structure of the homology model of *Tg*SPT as a holo-monomer with PLP bound in the active site. The PLP cofactor is shown as a grey stick model. The monomer is coloured with the central catalytic domain (residues 215 - 449) shown in pale green and the N-terminal domain and an α/β structure which contains a six-stranded β-sheet twisted in a spiral manner surrounded by three α-helices (residues 156 - 178, 178 - 214 and 450 - 548) in dark green. Figure was prepared using *PyMOL*.[192]

The monomer consists of three domains (only two are shown in the Figure 4.10 as the N-terminal has not been modelled). However, in the *Sm*SPT monomer the first 23 residues make up the N-terminal domain (equivalent to residue number 178 for *Tg*SPT) which contains two small α-helices. The residues from 24 to 59 (179 - 214 in *Tg*SPT) and 295 to 393 (450 - 548 in *Tg*SPT) make up two small parts of the sequence which form a small domain; the structure is an α/β structure containing a six-stranded β-sheet which is twisted in a spiral manner, and surrounded by three α-helices from the small domain outer surface side. The remaining residues make up the largest domain which assumes an α/β structure where a seven-stranded β-sheet and ten α-helices are twisted to form a large and open structure. The β-sheet is a characteristic of the PLP-enzymes belonging to the type I fold; six of the strands are parallel (strands number 3, 8, 7, 6, 4 and 5) with β9 being anti-parallel, positioned between β3 and β8.[176]

PLP was built into the homology model and a close-up of the resulting predicted PLP binding site is shown below in Figure 4.11.

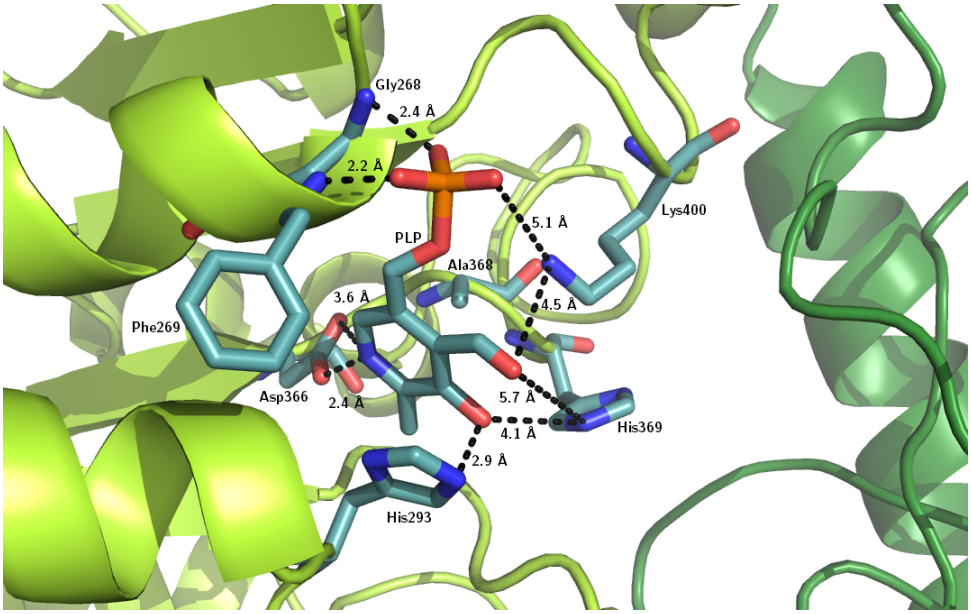


Figure 4.11: Close up of the active site of the homology model of *Tg*SPT with PLP bound in the active site. The PLP cofactor is shown as a grey stick model. The N-terminal and central catalytic domain are shown in pale green and the active site residues involved in binding the PLP cofactor are shown in teal and labelled. Relevant hydrogen bonds are shown with distances measured in Å. Residues Tyr209 and Ser399 as well as Ser429* and Ala430* have been omitted for clarity. * Denote the residues located on the other monomer. Figure was prepared using *PyMOL*.[192]

Figure 4.11 shows that the residues, Gly268, Phe269, His293, Lys400, Asp366 and His369 are all forming hydrogen bonds with the PLP molecule contained in the active site. The active site lysine, residue number 400, is also conserved for *Tg*SPT. The active site residues are positioned near the end of the β-sheet consisting of seven strands.[176] The PLP cofactor is positioned in the active site but is held in place by hydrogen bonds and hydrophobic interactions. Also, there is no covalent bond between the cofactor and the active site lysine (evidence from electron density for *Sm*SPT). The Lys400 side chain is not directed towards the C4 of PLP.
The PLP pyridinium ring is sandwiched between the His293 side chain and the Ala368 side chain stabilised by hydrophobic interactions and a hydrogen bond with Asp366 side chain to the N atom which has a $\delta+$ charge. The His293 imidazole ring lies parallel to the PLP pyridinium ring enabling the oxygen atom on the ring to form a hydrogen bond with Nϵ2 on His369.
Meanwhile, the phosphate headgroup of PLP is connected to the active site residues with polar interactions with the N atoms of the main chain in F269.

The N-terminal domain contains the transmembrane helix responsible for membrane anchoring plus a domain of unknown structure and function. Based on the homology model this domain is not involved in the enzymatic reaction.

As the bacterial orthologues all function as homodimers a similar homodimer was generated using Coot.[191] As shown in Figure 4.12 the homology model clearly allows a compact dimeric structure with residues from both subunits contributing to the active site in the other subunit.

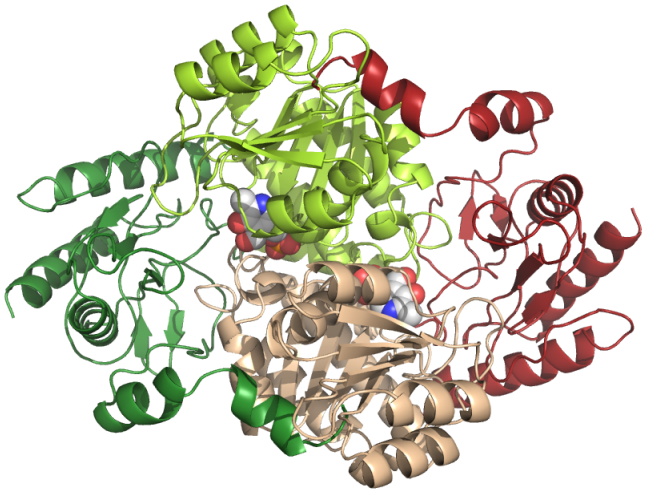


Figure 4.12: Homology model of the holo *Tg*SPT dimer. One monomer is coloured with the N-terminal and central catalytic domain in pale green and the C-terminal domain in dark green. The second subunit is coloured similarly in brown and red. The PLP cofactors are shown as grey spheres (with O atoms in red, N atoms in blue and P atoms in orange). Figure was prepared using *PyMOL*.[192]

The homodimer shows the two monomers interacting in a manner in which the active site is supplemented by residues from the opposite monomer resulting in better binding of the substrates.[176] The substrate binding pocket is located where the PLP molecule binds; this is situated between the two monomers and residues from the opposite monomer aid in securing the cofactor molecule in the opposing monomer. Residues from both subunits complement the active site suggesting that the enzyme is active as a dimer with a surface interface of 4701.8 Å^2 (calculated using the 'Protein interfaces, surfaces and assemblies' service PISA at the European Bioinformatics Institute, available at http://www.ebi.ac.uk/pdbe/prot_int/pistart.html).[193]

4.2.3 Cloning and small-scale expression trials at the OPPF

The initial cloning and small-scale expression experiments described in this section were undertaken by I. R. Williamson at the OPPF. Based on the initial results (Figure 4.14) the following four constructs (of the original 48) that showed the highest expression levels of soluble protein were selected for further analysis (Table 4.1).

Construct label	aaN	aaC	pOPIN vector	Tag	MW/ kDa	Plasmid name
Δ143	144	571	S3C	N-His-SUMO3C	47.91	*pSPT-E6*
Δ158	159	571	S3C	N-His-SUMO3C	45.98	*pSPT-F6*
Δ176	177	571	S3C	N-His-SUMO3C	43.86	*pSPT-G6*
Δ180	181	571	S3C	N-His-SUMO3C	43.29	*pSPT-H6*

Table 4.1: Summary of the four SPT constructs, including construct label, N- and C-terminal, pOPIN vector employed, construct size, molecular weight and plasmid label. The molecular weight represents the proteins without the tags. The His_6-SUMO tag has a mass of 13213 Da. All proteins contain a 3C cleavage site.

These constructs were all expressed from a pOPINS3C vector and contain an N-terminal His_6-SUMO tag. The hexa-histidine part enables IMAC purification and the SUMO (Small Ubiquitin-like Modifier) is mainly employed for solubility reasons.

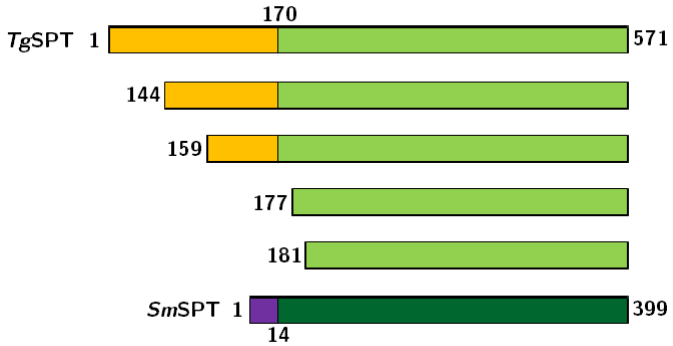
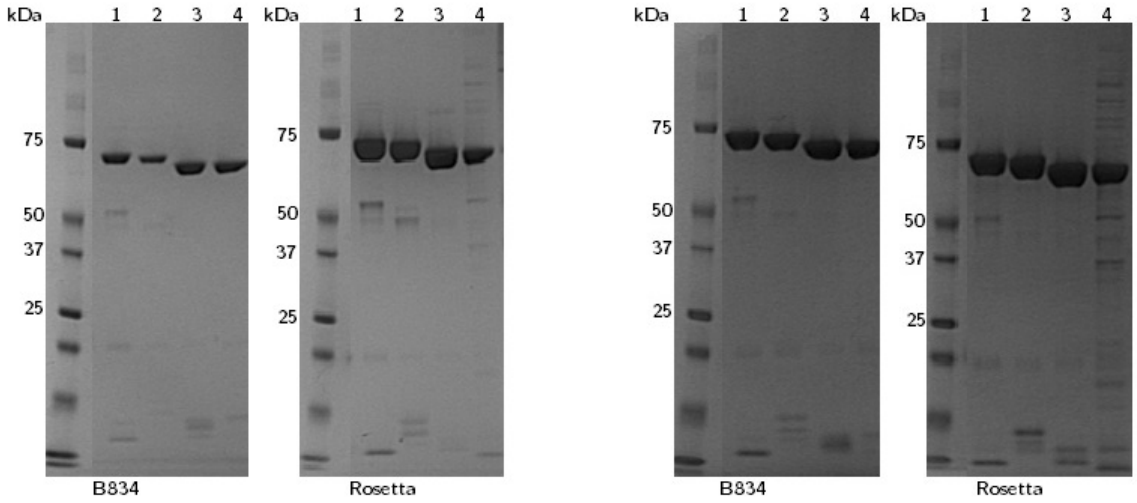


Figure 4.13: Schematic diagram showing the full length *Tg*SPT and the four constructs compared to the *Sm*SPT. The light green (*Tg*SPT) and dark green (*Sm*SPT) domains are the conserved domains of the SPTs and the orange (*Tg*SPT) and purple (*Sm*SPT) N-terminal domains which is membrane associated for *Tg*SPT. The numbers mark the relevant residue numbers.

Figure 4.13 shows a schematic representation of the constructs of *Tg*SPT cho-

sen for expression experiments compared to *Sm*SPT. The constructs employed in this chapter get sequentially shorter by removal of N-terminal residues. It is clear from the diagram that the shorter constructs possess none of the membrane associated domain of *Tg*SPT. Common to all constructs is the conserved domain which resembles the bacterial orthologues.

Small-scale expression trials performed at the OPPF (similar strategy was employed for the PlcH constructs, see 2.2.10) were carried out using B834 and Rosetta I (DE3)pLysS cells both using IPTG medium (IPTG induction) and Overnight Express^{TM} Instant TB medium (autoinduction). The induced cultures were centrifuged and the supernatant analysed on a 12% SDS-PAGE gel. The results for the constructs showing over-expression of soluble protein are shown in Figure 4.14 (partial results from data obtained at the OPPF by I. R. Williamson (I. Williamson, M. Phil thesis, Durham University, 2012)).



(a) (b)

Figure 4.14: 12% SDS PAGE of 5 μL of resolved, purified protein (for constructs SPT Δ143, SPT Δ158, SPT Δ176 and SPT Δ180). Proteins were expressed from two *E. coli* strains; B834 (left side of gels) and Rosetta I (DE3)pLysS. Expression was carried out using IPTG induction medium (a) and OvernightExpress^{TM} Instant TB autoinduction medium (b). Proteins were stained with Instant Blue (Expedeon) dye and protein size was estimated relative to a Precision Plus Protein^{TM} Standard (BioRad). These are extracts of figures of four gels obtained at the OPPF (I. Williamson, M. Phil thesis, Durham University, 2012).

Figure 4.14 shows increased bands for the four SPT constructs Δ143, Δ158,

Δ176 and Δ180 from both B834 cells and Rosetta I (DE3)pLysS cells when induced by IPTG as well as by autoinduction. More protein is expressed from Rosetta I (DE3)pLysS cells and when using autoinduction rather than IPTG induction.
Based on the positive results achieved at the OPPF, these four constructs were put forward to large-scale expression and purification described below in section 4.2.4.

4.2.4 Optimising protein expression of *Tg*SPT

Large-scale expression of the constructs *SPT Δ143, SPT Δ158, SPT Δ176* and *SPT Δ180* in Rosetta II (DE3)pLysS cells (Novagen) was carried out as described in section 2.4.3. The pure proteins were analysed on a 12% SDS-PAGE gel as shown in Figure 4.15.

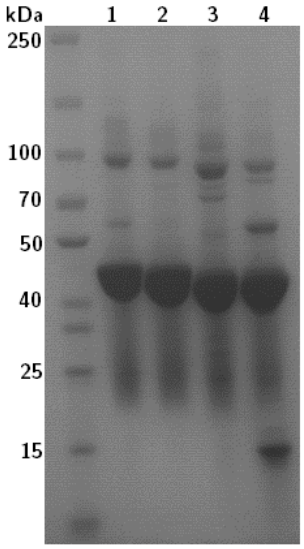


Figure 4.15: 12% SDS PAGE of ∼25 μmol of resolved, purified, cleaved protein in lanes labelled 1 - 4; SPT Δ143, SPT Δ158, SPT Δ176 and SPT Δ180 respectively. Proteins were stained with Instant Blue (Expedeon) dye and protein size was estimated relative to a Spectra Multicolour Broad Range Protein Ladder (Thermo Scientific).

The SDS-PAGE gel in Figure 4.15 shows large amounts of pure, cleaved protein for constructs SPT Δ143, Δ158, Δ176 and Δ180; it is clear from the bands that the proteins appear at the correct molecular mass (see Table 4.1) and that Δ143 is the largest construct and Δ180 the shortest.
The gel is overloaded with protein but this provides useful information. Bands

are present at twice the mass for each protein indicating that the SPT constructs form homo-dimers. The pure protein had a distinct yellow colour which was noticed during initial purifications, indicative of PLP being bound. Thus, PLP was added to all purification buffers and sample brightness increased upon increased availability of the PLP cofactor.
Although high expression levels were achieved, protein stability for the two shortest constructs remained low (ie. storage at 4 °C overnight for further purification and concentrating often resulted in protein precipitation by aggregation, especially for SPT Δ180). The lane containing SPT Δ143 also reveals more impurities than the other constructs. Typical yields were reported as 3 mg/L expression medium for Δ143, 32 mg/L expression medium for Δ158, 14 mg/L expression medium for Δ176 and 22 mg/L expression medium for Δ180. Sample purity was highest for the largest constructs, SPT Δ143 and SPT Δ158.

In addition, samples of SPT Δ158 (45.98 kDa) and SPT Δ180 (43.29 kDa) were applied to a native gel (section 2.5.7) to avoid denaturing the proteins and perform a preliminary investigation into potential dimerisation.

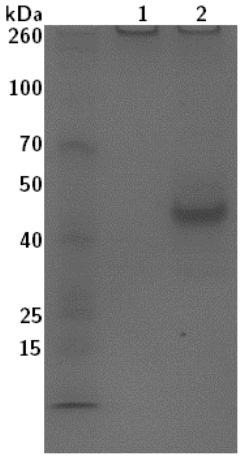


Figure 4.16: Native gel of purifed, cleaved protein in lanes 1: SPT Δ158 and 2: SPT Δ180. Proteins were stained with Instant Blue (Expedeon) dye and protein size was estimated relative to a Spectra Multicolour Broad Range Protein Ladder (Thermo Scientific).

The native gel in Figure 4.16 shows no bands for SPT Δ158 and only shows a single band at approximately 43 kDa for SPT Δ180. This suggests SPT Δ180 exists in a monomeric form under the conditions of the gel. SPT Δ158 may not have entered the gel which could be a result of precipitation. The results for SPT Δ158 are inconclusive, thus more experimental data are needed.

4.2.5 Protein mass spectrometry

The protein sample of SPT Δ158 was analysed by mass spectrometry (section 2.6.2). The data are shown below in Figure 4.17. The accurate, calculated mass of SPT Δ158 is 45980.8 Da (413 residues) and the M^+ peak in the spectrum collected on the QToF Premier is at 46132 Da ± 2, corresponding to a difference in mass of 150.2 Da. The tag was cleaved off the protein employing the HRV 3C protease (a cysteine protease) which recognises the specific cleavage site Leu-Glu-Val-Leu-Phe-Gln*Gly-Pro, leaving a glycine and a proline residue at the N-terminal of the protein. The two additional amino acids results in a total mass of 46134.9 Da, which is within the experimental error of the $[M^+]$ peak. The $2\times[M^+]$ peak is found at a mass of 92264 Da. This is a strong peak confirming the notion that *Tg*SPT does, at least partially, exist as a homo-dimer, as suggested by the bioinformatical information and the Bio-SAXS experiments (section 4.2.5.1).

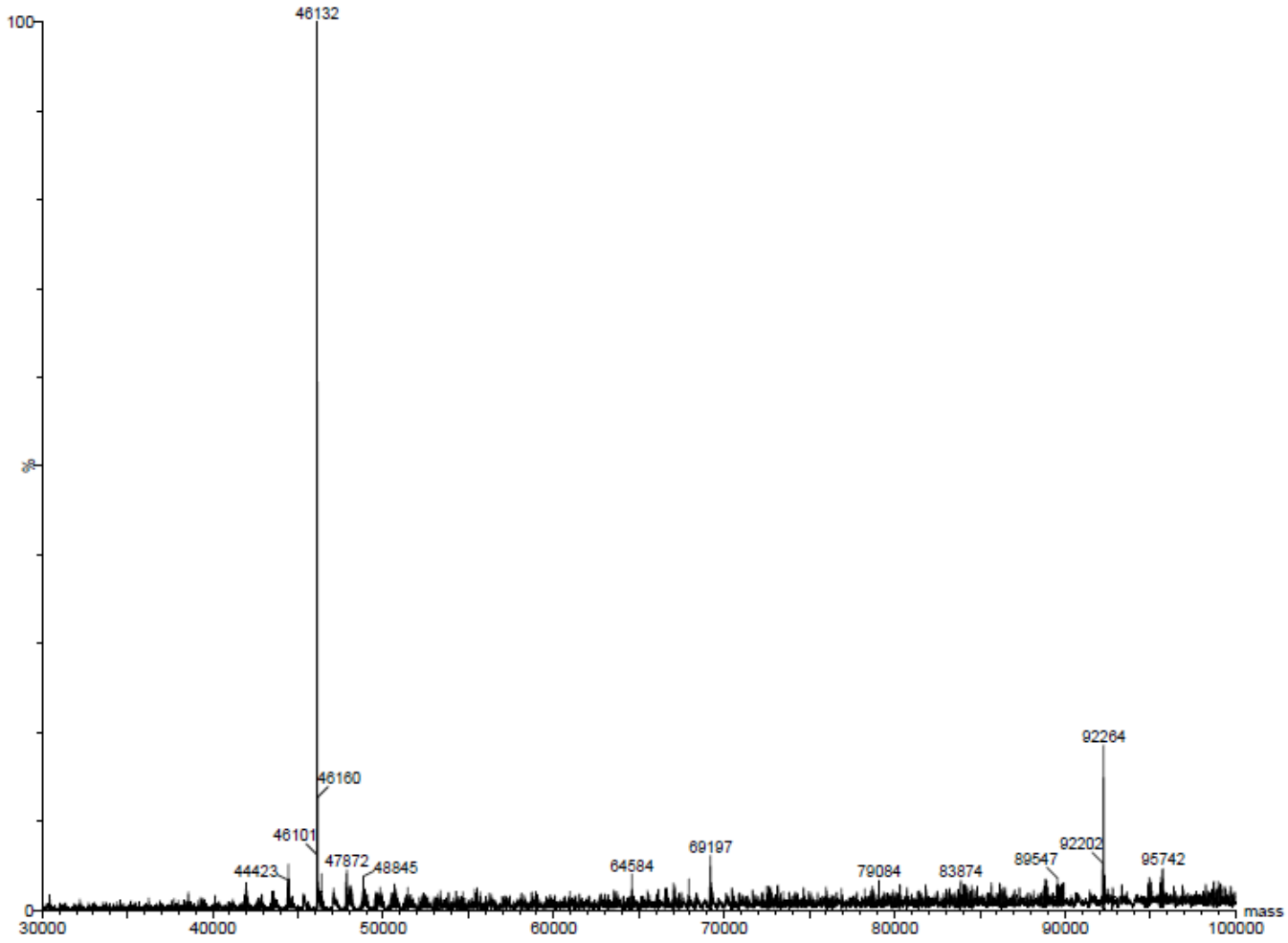


Figure 4.17: LC-ESI-MS of SPT Δ158 in 40 mM ammonium acetate pH 7, 25 μM PLP.

The data shows unambiguously that the correct protein, SPT Δ158, is expressed and that there are no undesireable post-translational modifications such as oxidations present in the protein. In addition, there is strong evidence that *Tg*SPT is expressed and purified as a homodimer. There is no evidence for binding of the PLP cofactor, even though the yellow colour of pure protein strongly suggests that the cofactor is bound. However as the protein is transformed to the gas phase, the cofactor can easily be lost as it is not tightly bound.

4.2.5.1 Small angle X-ray scattering experiments

Small angle X-ray scattering experiments were performed in order to determine the overall shape of the putative SPT homodimer. A representative plot of I * s^2 versus s (Kratky plot) and log(I) versus s (log-lin plot) are shown for the data collected for SPT Δ158 at a concentration of 3.8 mg/mL are shown below in Figures 4.18 and 4.19 respectively. In the figures, I is the scattered intensity and s is the momentum transfer.

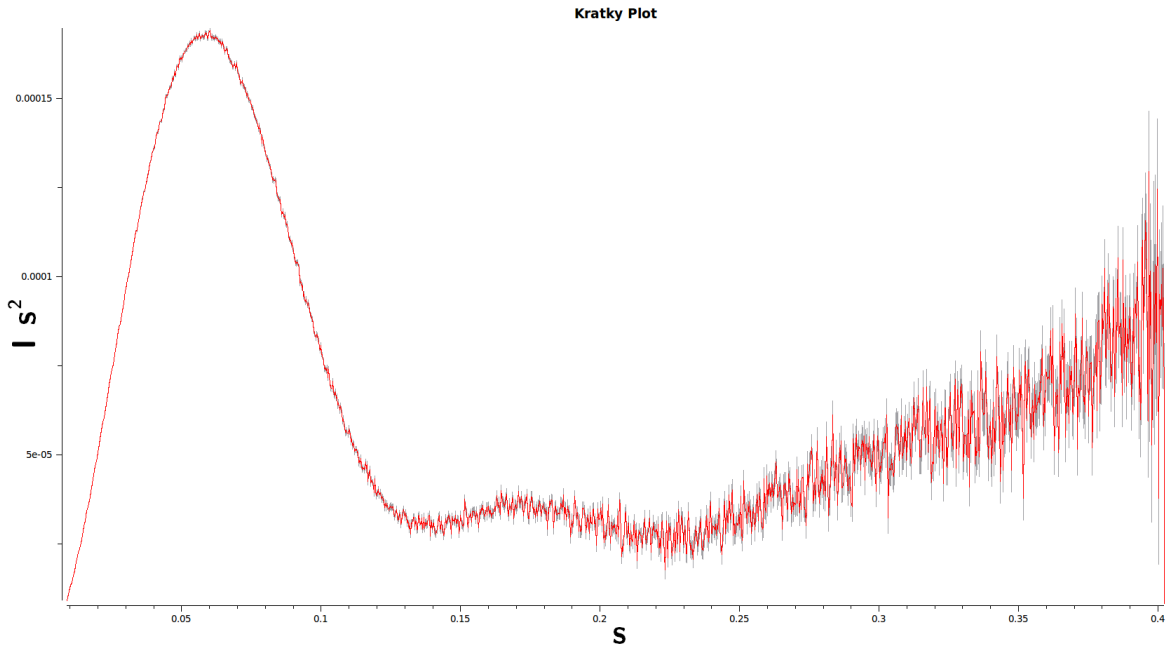


Figure 4.18: Kratky plot of SPT Δ158 at 3.80 mg/mL showing I * s^2 versus s (I is the intensity, s is the momentum transfer). Figure generated in SASPLOT from the ATSAS programme package.

The representative Kratky plot in Figure 4.18 of SPT Δ158 at 3.80 mg/mL has a scattering profile which indicates that the protein is folded. This is verified by the

decay profile, as the intensity usually decays as $1/s^4$ for folded proteins. A peak is usually present at low angles in the Kratky plot for folded, globular proteins.[194]

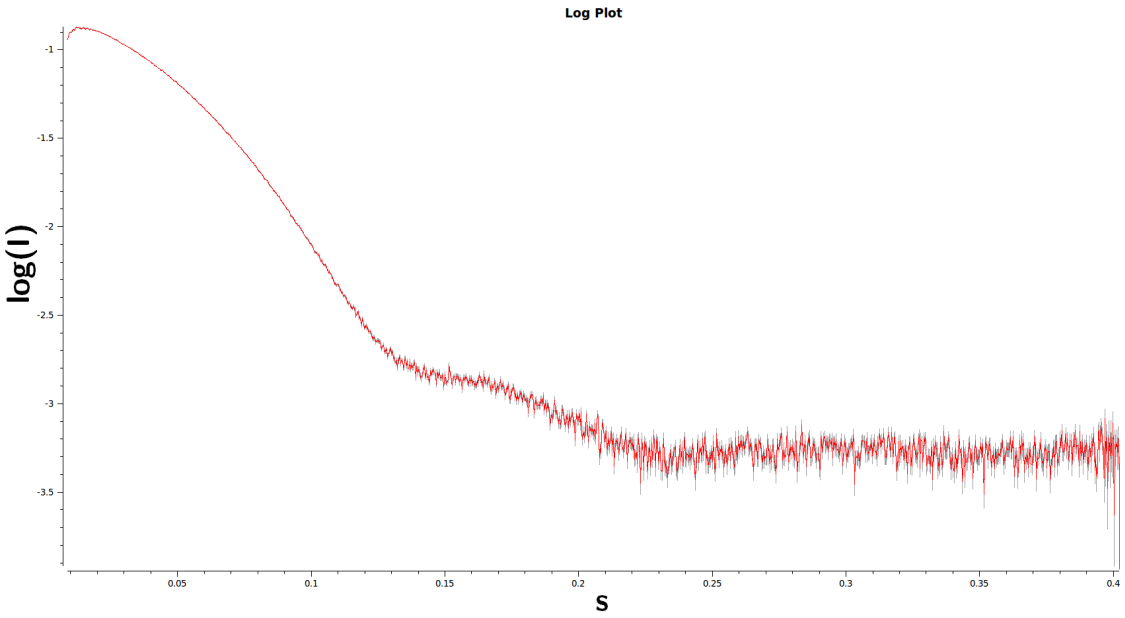


Figure 4.19: SAXS curve of SPT Δ158 at 3.84 mg/mL showing log (I(s)) versus s (I is the intensity, s is the momentum transfer). Figure generated in SASPLOT from the ATSAS programme package.

Figure 4.19 shows the log of the intensity plotted versus s. The shape of the curve fits with that expected for a folded protein in solution. At the very lowest angles there is a dip in intensity and this data were not included in further data analysis. The curves for all four SPT constructs were representative of homogeneous protein samples. Further analysis provided the approximate MW.

The analysed Bio-SAXS data are shown in Tables 4.2, 4.3, 4.4 and 4.5 for SPT Δ143, SPT Δ158, SPT Δ176 and SPT Δ180 respectively. Porod volumes were calculated in Primus from the ATSAS package and the Porod MW was calculated.

Concentration/ mg mL^{-1}	Porod Volume/ $Å^{-3}$	Porod MW/kDa	Expected MW/kDA
1.50	209161	130.7	47.9
1.20	186516	116.6	47.9
0.75	198277	123.9	47.9
0.15	184332	115.2	47.9
1.50	181724	113.6	47.9
1.20	209570	131.0	47.9
0.75	188557	117.8	47.9
0.15	197805	123.6	47.9

Table 4.2: Experimental and analysed Bio-SAXS data showing the sample concentration, the measured Porod volume, the estimated Porod molecular weight and the expected molecular weight for SPT Δ143.

Table 4.2 shows the data for SPT Δ143. Despite the data being close to the values expected for an SPT Δ143 dimer, the estimated values for the molecular weight do not correlate well with the expected values and are not consistent between measurements at different protein concentrations and different data collections. The data may indicate that SPT Δ143 aggregates in solution causing the discrepancy between the measured values and the predicted molecular weights.

Concentration/ mg mL^{-1}	Porod Volume/ $Å^{-3}$	Porod MW/kDa	Expected MW/kDA
4.80	145232	90.8	46.0
3.84	143570	89.7	46.0
2.40	137332	85.8	46.0
2.40	135714	84.8	46.0
4.80	141760	88.6	46.0
3.84	141862	88.7	46.0
2.40	136186	85.1	46.0
2.40	125169	78.2	46.0

Table 4.3: Experimental and analysed Bio-SAXS data showing the sample concentration, the measured Porod volume, the estimated Porod molecular weight and the expected molecular weight for SPT Δ158.

Table 4.3 shows the data for SPT Δ158. The numerical data shows excellent correlation between the estimated molecular weight based on the SAXS data and the expected molecular weight of the protein, especially at higher protein concentrations. The data are consistent with SPT Δ158 forming homodimers in solution.

Concentration/ mg mL^{-1}	Porod Volume/ Å^{-3}	Porod MW/kDa	Expected MW/kDA
4.60	137310	85.8	43.9
3.68	131613	82.3	43.9
2.30	125499	78.4	43.9
0.46	122996	76.9	43.9
4.60	132375	82.7	43.9
3.68	129764	81.1	43.9
2.30	125197	78.2	43.9
0.46	133333	83.3	43.9

Table 4.4: Experimental and analysed Bio-SAXS data showing the sample concentration, the measured Porod volume, the estimated Porod molecular weight and the expected molecular weight for SPT Δ176.

Table 4.4 shows the data for SPT Δ176. As for SPT Δ158, the numerical data shows excellent correlation between the estimated molecular weight based on the SAXS data and the expected molecular weight of the protein, especially at higher protein concentrations. The data are consistent with SPT Δ176 forming homodimers in solution.

Concentration/ mg mL^{-1}	Porod Volume/ Å^{-3}	Porod MW/kDa	Expected MW/kDA
2.16	112464	70.3	43.3
1.73	111622	69.8	43.3
1.08	111353	69.6	43.3
0.22	121630	76.0	43.3
2.16	112537	70.3	43.3
1.73	111390	69.6	43.3
1.08	113194	70.7	43.3
0.22	124426	77.8	43.3

Table 4.5: Experimental and analysed Bio-SAXS data showing the sample concentration, the measured Porod volume, the estimated Porod molecular weight and the expected molecular weight for SPT Δ180.

Table 4.5 shows the data for SPT Δ180. Despite the data being close to the values expected for an SPT Δ180 dimer, the estimated values for the molecular weight do not correlate well with the expected values. They are not consistent between measurements at different protein concentrations, especially lower protein concentrations. The data may indicate that SPT Δ180 is transient between a monomer

and a homodimer in solution causing the discrepancy between the measured values and the predicted molecular weights.

Overall, the data clearly indicate that the SPT Δ158 and Δ176 construct form homodimers in solution. The two other constructs, SPT Δ143 and SPT Δ180, probably also form homodimers, however, conditions may have caused them to either aggregate or adapt a transient state between monomer and dimer in solution.
In order to reach a clear conclusion, further analysis needs to be performed to calculate a molecular envelope in order to ascertain whether all the SPT constructs form homodimers in solution.

4.2.6 CD spectroscopy on SPT constructs

In order to verify the correct folding of all protein constructs, CD spectra were recorded as described in section 2.7.
The experimental, analysed spectra are shown in Figure 4.20.

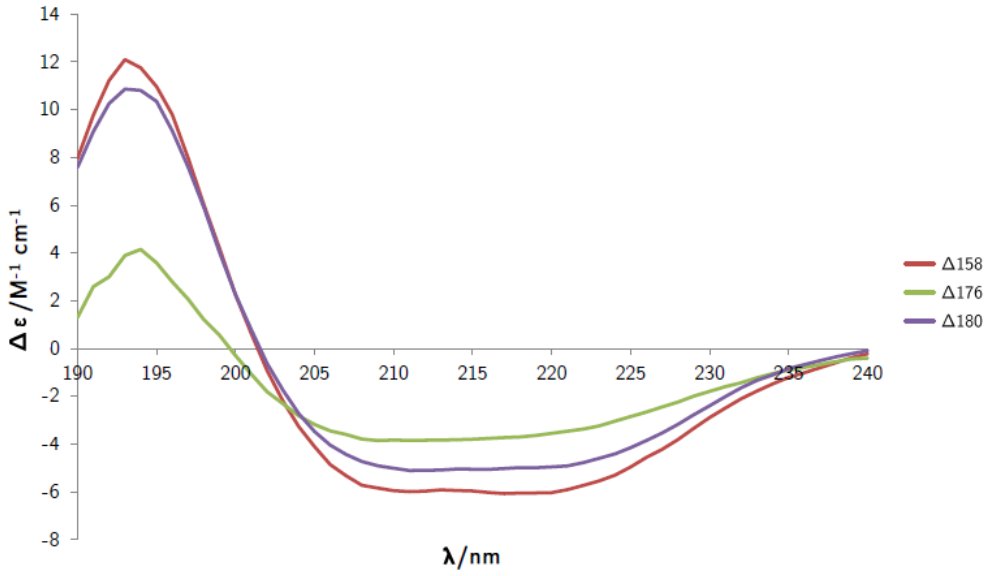


Figure 4.20: Experimental CD data, molar circular dichroism, plotted against wavelength (nm) for the three shortest SPT constructs; SPT Δ158, SPT Δ176 and SPT Δ180 (0.18 - 0.25 μM) measured at 21 °C.

Figure 4.20 shows the experimental spectra, $\Delta\,\epsilon$ plotted against the wavelength. Qualitatively, all three spectra show the characteristic shape of a folded protein with a significant proportion of α-helices.

The data were analysed using three different programmes - CONTINLL, CDSSTR and SELCON3 - in order to estimate the secondary structure. The analysis yielded molar circular dichroism, $\Delta\ \epsilon$, calculated by each programme (using the same basis set of proteins) which was then compared to the experimentally measured $\Delta\ \epsilon$ values (Appendix C, Figures C.1, C.2 and C.3). For the three constructs, SPT Δ158, SPT Δ176 and SPT Δ180, the CONTINLL programme analysis calculating an experimental CD spectrum most closely matched the measured, experimental data, closely followed by the CDSSTR programme. The percentages of secondary structure elements as calculated by the programmes is shown in Table 4.6.

Construct	α-helices (%)	β-sheets (%)	Turns (%)	Unordered (%)
Δ158	55	8	14	23
Δ176	33	19	24	24
Δ180	53	9	14	24

Table 4.6: Calculated secondary structure elements for three SPT constructs. Data from a CONTINLL analysis showed for SPTs Δ158, Δ176 and Δ180.

The more quantitative analysis was performed by the CONTINLL, CDSSTR and SELCON3 programmes calculated the distribution of secondary structure elements of the three SPT constructs which is summarised in Table 4.6. The three SPT constructs all consist mainly of α-helices; the percentage of α-helices decreases as the construct size does. The percentage of turns increases as the construct size decreases. SPT Δ176 does contains a larger number of turns which may be due to poorer sample quality as a result of freezing.
The CD spectroscopy data shows that all the expressed and purified constructs for *Tg*SPT are properly folded proteins as the spectra display features characteristic of folded proteins consisting mainly of α-helices.

4.2.7 [C^{14}]-serine activity assay on *Tg*SPT constructs

Once the protein constructs were shown to be correctly folded, the enzymatic activities were determined.
The activity of the four SPT constructs was tested using an *in vitro* assay; [C^{14}]-serine and palmitoyl-CoA were employed as substrates to test for SPT functionality as described in 2.9.3. The results of the fluorographic analysis are shown below in Figure 4.21.

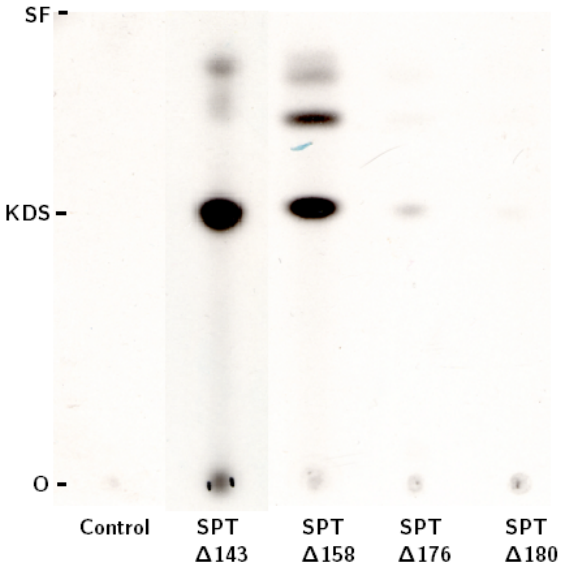


Figure 4.21: Fluorographic imaging of TLC plate blotted with 8 μL of product from SPT condensation of [C^{14}]-serine and palmitoyl-CoA for the four SPT constructs. The product KDS is found at an R_f value of $\sim$ 0.6. The control was the assay performed without the addition of SPT.

As seen in Figure 4.21 a major spot is detectable at an R_f value of $\sim$ 0.6; as the SPT constructs get shorter, the spots get fainter, indicating that longer constructs possess a higher enzymatic activity. The strong bands seen at R_f values of $\sim$ 0.6 corresponds to the product of SPT activity on palmitoyl-CoA and [C^{14}]-serine, KDS. This was verified by utilising *Sp*SPT as control for known SPT activity; products were observed at the same R_f value for *Sp*SPT and the four *Tg*SPT constructs (results omitted). The location of the catalysis product, KDS, was identified on the TLC plate by comparison with *Sp*SPT expressed in *E. coli* as a control (results not shown). The spots appearing at larger R_f values of $\sim$ 0.9 are a secondary reaction product that has not yet been identified.
The four SPT constructs all possess SPT activity *in vitro*; however, the two shorter constructs, SPT Δ176 and SPT Δ180, are significantly less active as evidenced by the faint spots for the product, KDS (Figure 4.21). The SPT activity assay was employed to identify the minimally catalytically active domain. Full activity was reached only with the longest protein construct, SPT Δ143.

4.2.7.1 Condensation product verification by mass spectrometry

The identity of the major product formed during the condensation reaction, KDS ($C_{18}H_{37}NO_2$, MW = 299 g/mol), was confirmed by mass spectrometry. The reaction was performed identically to the activity assay only the [C^{14}]-serine was replaced by cold L-serine. SPT Δ158 was used. This experiment had previously been performed with Δ143 but due to the stronger, more defined spots for the secondary product observed in Figure 4.21 for Δ158, this particular construct was chosen in order to attempt a characterisation of the secondary product, also.
Analysis was carried out by QToF Premier XE (Waters Ltd.) and the accurate mass spectrum is shown in Figure 4.22 below.

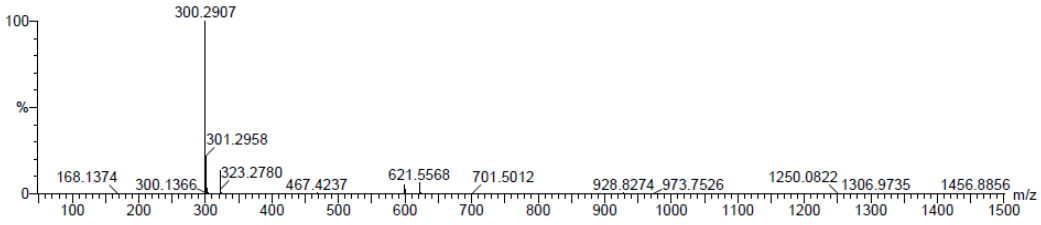


Figure 4.22: Mass spectrum of lipid fraction extracted after SPT condensation assay showing a major $[M+H]^+$ peak at 300.2907 *m/z* with the proposed molecular formula $C_{18}H_{38}NO_2$ (KDS).

The spectrum in Figure 4.22 shows an $[M+H]^+$ peak at 300.3 *m/z* at 100% abundance. The proposed molecular formula is given as $C_{18}H_{38}NO_2$, corresponding to the $[M+H]^+$ ion (quoted as 1.3 ppm). It is noteworthy that no secondary product is identifiable from the mass spectrum. This result unambiguously confirms the identity of the condensation reaction product as KDS.

4.2.7.2 SPT activity inhibition by myriocin

Myriocin (Figure 4.8) is a natural fungal product which is a known inhibitor of SPT activity (bacterial, yeast as well as mammalian orthologues). It is known to strongly bind the bacterial enzyme and was therefore chosen to be tested for inhibitory effects on the recombinant *Tg*SPTs as a first step of a later stage of drug design.
The assay was performed as described in section 2.9.3 and myriocin was added to

a final concentration of 100 μM (or CH_3OH) and the mixtures were incubated for 60 minutess before the palmitoyl-CoA and [C^{14}]-serine was added to the reaction mixture.

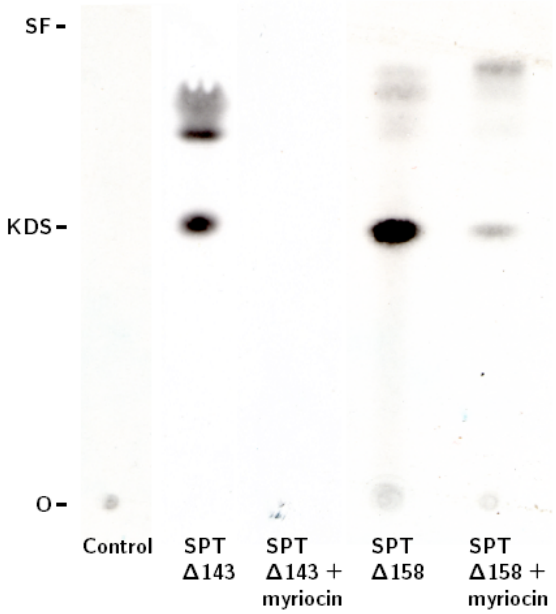


Figure 4.23: Fluorographic imaging of TLC plate blotted with 8 μL of product from SPT condensation of [C^{14}]-serine and palmitoyl-CoA for SPT Δ143, SPT Δ143 with 100 μM myriocin, SPT Δ158 and SPT Δ158 with 100 μM of myriocin. The product KDS is found at an R_f value of $\sim$ 0.6. The control was the assay performed without the addition of SPT

Figure 4.23 shows strong spots for both SPT Δ143 and SPT Δ158, producing the radiolabelled product, [C^{14}]-KDS (as marked on the TLC plate). When adding myriocin, the spot completely disappears when using SPT Δ143, whereas for SPT Δ158, a faint spot is still visible for [C^{14}]-KDS.
Myriocin completely inhibits the SPT activity of SPT Δ143 at a concentration of 100 μM, whereas only partial inhibition is detected for SPT Δ158. It also inhibits the members of the large and diverse SPT family, suggesting that the reaction mechanism remains conserved in the family.

4.2.7.3 Characterisation of SPT inhibitor myriocin

To determine the IC_{50} value of myriocin on SPT Δ158 at 20 μM, the procedure outlined in 2.9.3 was followed, except reaction volumes were reduced to 125 μL and myriocin concentrations ranging from 0 - 100 μM were added to the reaction mixture. The TLC plates were scanned using an AR-2000 radio TLC Imaging

Scanner (Bioscan). A scan of each lane produced a spectrum of radioactive counts and using the Laura software (LabLogic), the area under the peaks was determined. The background measurements, of reaction mixture containing no SPT protein, were subtracted from all other measurements. Figure 4.24 shows the radioactive counts (averaged over 4 independent measurements) measured (across each 10 cm long lane over 3 minutes) versus the myriocin concentration.

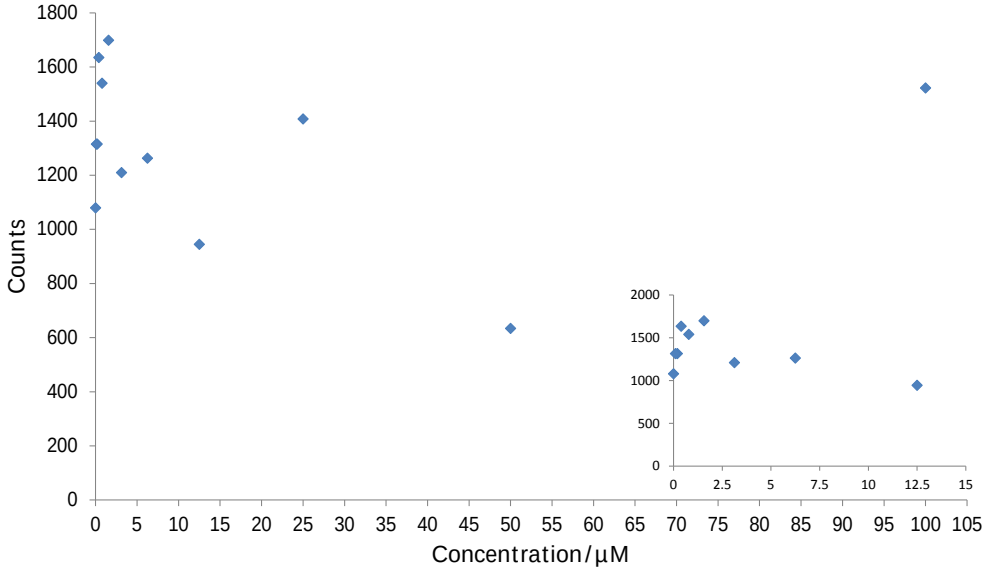


Figure 4.24: Inhibition data of SPT Δ158 by myriocin. Data obtained at a constant concentration of SPT Δ158 of 20 μM with concentrations of myriocin between 0 and 100 μM. The number of radioactive counts detected (normalised) on the TLC plate for the [C^{14}]-KDS is plotted against the myriocin concentration. Data based on four independent measurements. Insert shows a the data points at myriocin concentrations between 0 and 12.5 μM. Error bars emitted for clarity.

The data in Figure 4.24 does not show a clear trend and it is thus not possible to draw any clear conclusions from this data or calculate any IC_{50} values, and further experiments need to be conducted.

4.2.8 Thermal stability assay with *Tg*SPT

The goal of the thermal stability assays (TSAs) is to identify conditions that increase the thermal stability of the soluble protein sample. Employing medium-throughput methods developed in the laboratory allows for the identification of conditions and ligands that have the potential to increase the likelihood of crystallisation when used as additives.[54] The aim was to vary salt concentration, ionic strength or buffer type as much as possible with the various additives.
A graph showing the difference in melting temperature between the negative control (protein sample and dye with no additives) and conditions containing additives, ΔT, was plotted versus the additive concentration (linked to ionic strength). In Figure 4.25, this data has been shown for the compounds Na_2SO_4, $(NH_4)_2SO_4$, $Na_2C_3H_2O_4$ (disodium malonate) and $Na_3C_6H_5O_7$ (trisodium citrate) and NH_4Cl for concentrations ranging from 0.2 M to 1.5 M.

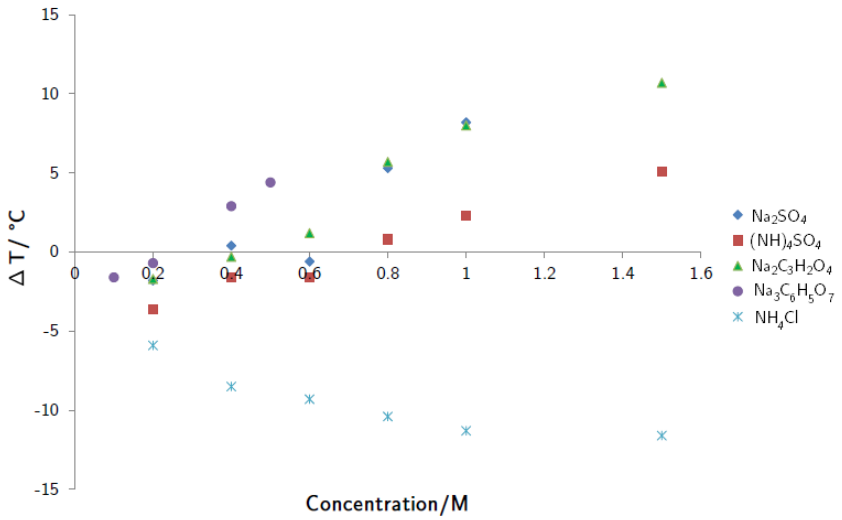


Figure 4.25: Graph showing the measured melting temperature change, ΔT, observed for SPT Δ158 at a final concentration of 0.8 mg/mL, 18 μM, plotted versus the concentration of the additive solution. Data from single measurement.

Figure 4.25 shows the change in protein melting temperature achieved through addition of varying concentrations of additives. The four compounds, Na_2SO_4, $(NH_4)_2SO_4$, $Na_2C_3H_2O_4$ and $Na_3C_6H_5O_7$ were found to have a stabilising effect on SPT Δ158 (an increase in T_{melt} is observed when additive is present) whilst NH_4Cl has a destabilising effect on SPT Δ158 (a decrease in T_{melt} is observed when

additive is present).
All the results are shown in Tables 4.7 and 4.8, which also contain information about additive solutions, concentrations, ionic strength and change in melting temperature of SPT Δ158 kept at a constant concentration of 0.8 mg/mL.
The data show that, within the measured concentration range, $Na_3C_6H_5O_7$ appears to increase the melting temperature of SPT Δ158 the most. However, larger concentrations of $(NH_4)_2SO_4$ and $Na_2C_3H_2O_4$ were tested; $Na_2C_3H_2O_4$ gave rise to the largest increase in stability of 10.7 °C at a 1.5 M concentration.

By plotting the measured temperature change versus the ionic strength (see Figure 4.26), the effect of ionic strength on melting temperature of the protein becomes apparent compared to the variations in ions employed and associated effect on protein interactions. Additives with high ionic strength tend to have a more stabilising effect on SPT Δ158.[195] Na_2SO_4 has the increase in T_{melt} with ionic strength whereas $(NH_4)_2SO_4$ has the smallest increase in T_{melt} with ionic strength (out of the stabilising compounds). $Na_2C_3H_2O_4$ and $Na_3C_6H_5O_7$ are second and third.

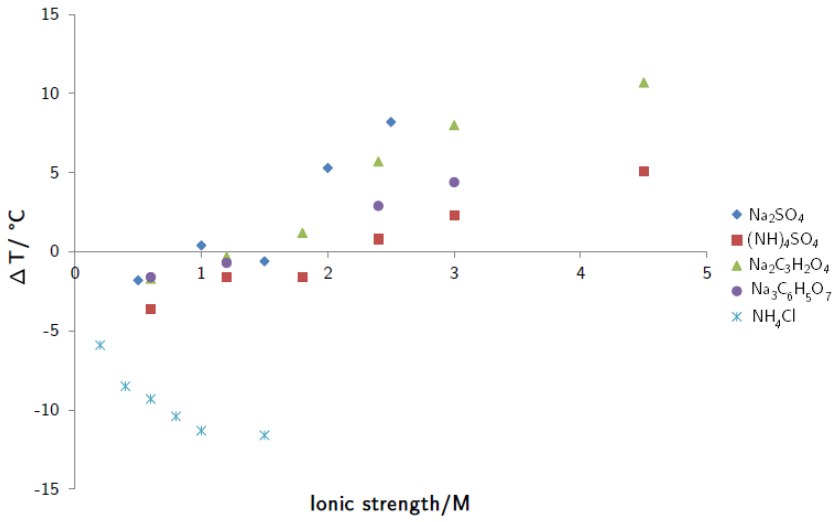


Figure 4.26: Graph showing the measured melting temperature change, ΔT, observed for SPT Δ158 at 18 μM, plotted versus the ionic strength of the additive solution. Data from single measurement.

The results in Figure 4.26 clearly indicate that the ionic strength plays an important role in the stabilising effect.

Compound	Concentration/M	T_{melt}/ °C	ΔT_{melt}/ °C	Ionic strength/M
Control (H_2O)		56.6		
TCA	7.5%	Protein denatured		
$MgSO_4$	1.0	57.0	0.4	4
$MgSO_4$	0.8	55.0	- 1.6	3.2
$MgSO_4$	0.6	54.0	- 2.6	2.4
$MgSO_4$	0.4	52.0	- 4.6	1.6
$MgSO_4$	0.2	51.0	- 5.6	0.8
Na_2SO_4	1.0	64.8	8.2	2.5
Na_2SO_4	0.8	61.9	5.3	2
Na_2SO_4	0.6	56.0	-0.6	1.5
Na_2SO_4	0.4	57.0	0.4	1
Na_2SO_4	0.2	54.8	-1.8	0.5
$(NH_4)_2SO_4$	1.5	61.7	5.1	4.5
$(NH_4)_2SO_4$	1.0	58.9	2.3	3
$(NH_4)_2SO_4$	0.8	57.4	0.8	2.4
$(NH_4)_2SO_4$	0.6	55.0	- 1.6	1.8
$(NH_4)_2SO_4$	0.4	55.0	- 1.6	1.2
$(NH_4)_2SO_4$	0.2	53.0	- 3.6	0.6
NaCl	1.5	52.0	- 4.6	1.5
NaCl	1.0	52.0	- 4.6	1
NaCl	0.8	51.0	- 5.6	0.8
NaCl	0.6	51.0	- 5.6	0.6
NaCl	0.4	51.0	- 5.6	0.4
NaCl	0.2	52.0	- 4.6	0.2
NH_4Cl	1.5	45.0	-11.5	1.5
NH_4Cl	1.0	45.3	-11.3	1
NH_4Cl	0.8	46.2	-10.4	0.8
NH_4Cl	0.6	47.3	-9.3	0.6
NH_4Cl	0.4	48.1	-8.5	0.4
NH_4Cl	0.2	50.7	-5.9	0.2
Na_2malonate	1.5	67.3	10.7	4.5
Na_2malonate	1.0	64.6	8	3
Na_2malonate	0.8	62.3	5.7	2.4
Na_2malonate	0.6	57.8	1.2	1.8
Na_2malonate	0.4	56.3	-0.3	1.2
Na_2malonate	0.2	54.9	-1.7	0.6
Na_3citrate	0.5	61.0	4.4	3
Na_3citrate	0.4	59.5	2.9	2.4
Na_3citrate	0.2	55.9	-0.7	1.2
Na_3citrate	0.1	55.0	- 1.6	0.6
KCl	0.5	51.0	- 5.6	0.5
KCl	0.2	52.0	- 4.6	0.2
LiCl	0.5	48.9	-7.7	0.5
LiCl	0.2	52.0	- 4.6	0.2
RbCl	0.5	50.6	- 6	0.5
RbCl	0.2	52.3	- 4.3	0.2
EDTA pH 8	0.005	56.9	0.3	
EGTA pH 8.9	0.005	57.1	0.5	

Table 4.7: Summary of the additives tested, their concentration, effect on melting temperature of SPT Δ158 at 18 μM and ionic strength. The screen tests various buffer, salt and ionic strength conditions.

Compound	Concentration/M	T_{melt}/ °C	ΔT_{melt}/ °C	Ionic strength/M
$MgCl_2$	0.005	56.0	-0.6	0.005
$CaCl_2$	0.005	55.5	-1.1	0.005
$SrCl_2$	0.005	55.6	-1	0.005
$MnCl_2$	0.001	55.5	-1.1	0.001
$ZnCl_2$	0.001	44.3	-12.3	0.001
$NiCl_2$	0.001	50.4	-6.2	0.001
$FeSO_4$	0.001	46.5	-10.1	0.004
$FeCl_3$	0.001	45.0	-11.6	0.006
$CoCl_2$	0.001	Protein denatured		0.003
$CuCl_2$	0.001	33.7	-22.9	0.003
$CdSO_4$	0.001	46.7	-9.9	0.004
$MgCl_2$, $CaCl_2$, $ZnCl_2$	all 0.001	43.5	-13.1	0.009
NaF	0.4	57.2	0.6	0.4
NaF	0.1	56.2	-0.4	0.1
NaBr	0.4	47.3	-9.3	0.4
NaBr	0.1	52.7	-9.3	0.1
NaI	0.4	37.5	-19.1	0.4
NaI	0.1	49.9	-6.7	0.1
Na formate	0.5	55.0	-1.6	0.5
Na formate	0.2	56.3	-0.3	0.2
Na acetate	0.5	54.9	-1.7	0.5
Na acetate	0.2	55.1	-1.5	0.2
Na L-lactate	0.5	57.8	1.2	0.5
Na L-lactate	0.2	57.3	0.7	0.2
$NaNO_3$	0.5	46.3	-10.3	0.5
$NaNO_3$	0.2	49.9	-6.7	0.2
Na_2HPO_4	0.005	57.0	0.4	0.03
Na_3VO_4	0.005	49.0	-7.6	0.03
Na_3WO_4	0.005	57.0	0.4	0.03
Na_2MoO_4	0.005	57.0	0.4	0.015
Urea	4	29.7	-26.9	
Urea	0.5	52.0	- 4.6	
Glycerol	10% (V/V)	57.2	0.6	
Glycerol	20% (V/V)	57.7	1.1	
CsCl	0.5	49.9	- 6.7	0.5
CsCl	0.2	51.5	-5.1	0.2
Bicine pH 8.5	0.1	52.0	- 4.6	
Tris-HCl pH 8.5	0.1	47.9	-8.7	
Tricine pH 8	0.1	55.0	-1.6	
PIPES pH 7.5	0.1	55.1	-1.5	
HEPES pH 7.5	0.1	55.3	-1.3	
MOPS pH 7	0.1	55.9	-0.7	
phosphate pH 7	0.1	53.0	- 3.6	
BisTrus propane pH 6.8	0.1	49.2	-7.4	
BisTris methane pH 6.75	0.1	54.0	-2.6	
MES pH 6.75	0.1	55.8	-0.8	
Acetic acid-NaOH pH 5	0.1	32.8	-23.8	
Na_3 citrate-HCl pH 5	0.1	43.8	-12.8	

Table 4.8: Summary of the additives tested, their concentration, effect on melting temperature of SPT Δ158 at 18 μM and ionic strength. The screen tests various buffer, salt and ionic strength conditions.

A commercial screen, the Hampton research Solubility and Stability HT Screen

(HR2-072) was also used to test for stabilising additives for SPT Δ158 (Table 4.9). Due to the large destabilising effect measured for most of the additives contained in this screen, it was not used for crystallisation trials.

Compound	Concentration/M	T_{melt}/ °C	ΔT_{melt}/ °C
Water		57.4	
Betaine monohydrate	2.5	61.8	4.4
Xylitol	2	61.2	3.8
D-Sorbitol	2	61.8	4.4
Sucrose	2	62.4	5
Trimethylamine N-oxide dihydrate	2.5	65.9	8.5
Sodium phosphate monobasic monohydrate,	0.7	63.3	5.9
Potassium phosphate dibasic	1.3		
L-Arginine ethyl ester dihydrochloride	0.5	33.8	-23.6
L-Argininamide dihydrochloride	0.5	33.8	-23.6
Spermidine	0.5	26.7	-30.7
Adipic acid	0.08	27.3	-30.1
Ethylenediamine dihydrochloride	0.5	33.3	-24.1
N-Methylformamide	30% (w/V)	31.3	-26.1
GSH (L-Glutathione reduced),	0.01	32.1	-25.3
GSSG (L-Glutathione oxidized)	0.01		
Ethylenediaminetetraacetic acid disodium salt dihydrate	0.1	29.0	-28.4
Cadmium chloride hydrate,	0.1	28.3	-29.1
Cobalt(II) chloride hexahydrate	0.1		
Non Detergent Sulfobetaine 201	1	26.3	-31.1
Ethylammonium nitrate	25% (w/V)	32.8	-24.6
Gadolinium(III) chloride hexahydrate	0.25	32.7	-24.7
Polyethylene glycol monomethyl ether 750	5% (w/V)	33.4	-24
Formamide	50% (V/V)	28.2	-29.2

Table 4.9: Summary of the additives tested, their concentration, effect on melting temperature of SPT Δ158 at 18 μM. The screen tests various salt, sugars and amino acid conditions

4.2.9 Crystallisation experiments of the SPT constructs

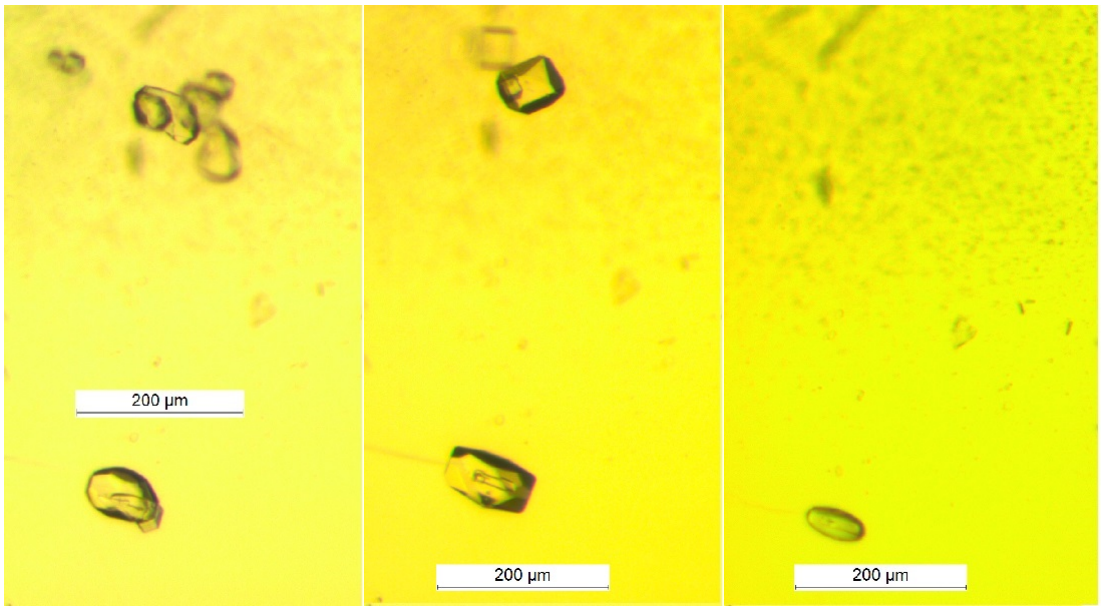
Crystallisation trays were set up for all four SPT constructs (section 2.9.6) at various concentration using the available crystallisation screens. After setting up crystallisation trays using all the available screens, none of the four constructs yielded any crystals. Additives were then added to the trays in an attempt to stabilise the enzymes.

For Δ143, 4200 conditions were tested up using different concentration of the protein. Myriocin, $Na_2C_3H_2O_4$ and $MgCl_2$ was added to some screens as an attempt to stabilise the protein. 2500 conditions were tested for Δ158, 1400 were tested for Δ176 and 2200 were tested for Δ180.

Most drops formed dark, brown amorphous precipitate however some were clear or

showed phase separation. Brown precipitate is often linked to denatured protein indicating that the protein is not particularly stable over the long time course of crystallisation in these conditions.

No crystalline conditions were found except for in the 96-well HTP tray of the Structure Screen 1+2[59,60] (Molecular Dimensions) for SPT Δ176 at 4.2 mg/mL. A cluster of oblong octagons were found in condition E8 (0.2 M ammonium phosphate monobasic, 0.1 M Tris pH 8.5, 50% (V/V) MPD) with a protein : reservoir ratio of 1 : 1. A month later, one of the crystals had grown larger (rectangular shape) and the others had merged into two larger crystals (rectangular and square). However, a week after that, the two smaller crystals had disappeared and the largest one shrunk to half its original size.



(a) (b) (c)

Figure 4.27: SPT Δ176 in 0.2 M ammonium phosphate monobasic, 0.1 M Tris pH 8.5, 50% (V/V) MPD (E6, Structure Screen 1+2)[59,60]. a) Oblong octagon shaped crystals. b) More defined oblong octagon shaped crystals one month after (a). c) Disintegrated crystals one week after (b).

The crystals were not tested in diffraction experiments as they deteriorated before they could be harvested. However, the decomposition pattern occurring over a short time-scale indicates that they could potentially be protein crystals.

4.3 Discussion

4.3.1 Discussion of the SPT homology modelling

Building a model using an enzymatic homologue is a useful way of predicting enzymatic function and comparing it to known protein structures. Even though *Tg*SPT and *Sm*SPT only possess a 27.8% sequence identity, resulting in only the conserved region being including in the model, it is still a good indicator of whether an enzymatic mechanism is conserved or not. Despite the low sequence identity, the active site residues surrounding the PLP cofactor still remain unchanged in *Tg*SPT, giving an indication that its active site and PLP binding site are very similar to that of *Sm*SPT. As the relevant residues are generally conserved, it strongly indicates identical, conserved reaction mechanisms.
Before expression and purification of the various *Tg*SPT constructs, the hypothesis was that the enzyme was functional as a homodimer. Construction of the homology model of *Tg*SPT showed a conserved active site around the PLP cofactor combined with interactions across a dimer interface, binding PLP with residues from each dimer subunit, indicating that *Tg*SPT is functional as a dimer. This is evolutionarily sound as higher organisms have developed towards heterodimers with similar monomer subunits having both evolved from the same precursor homodimer.

4.3.1.1 SPT expression

The small-scale over-expression experiments performed at the OPPF showed that the best results for expressing the highest amount of pure protein were obtained using Rosetta I (DE3) pLysS cells and the autoinduction method. This is probably due to the fact that Rosetta^{TM} strains are derivatives of BL21 cells; they have been modified to improve expression of eukaryotic proteins determined by specific codons rarely employed by *E. coli*. This strain provides higher levels of tRNA which recognises the mRNA linked to six extra DNA codons; AGA, AGG, AUA, CCC, CUA and GGA. This ensures that the protein expression is no longer limited by the *E. coli* codon usage (universal translation) and stabilises the recombinant vectors (pET vectors, which the pOPIN are a derivative of). The codon usage for *Tg*SPT can be compared to that of *E. coli* utilising the gcua (graphical codon usage analyser) software.[196] The results are shown in Figure 4.28, where the red bars represent the relative usage within the *Tg*SPT gene compared to the relative usage within *E. coli* represented by the black bars.

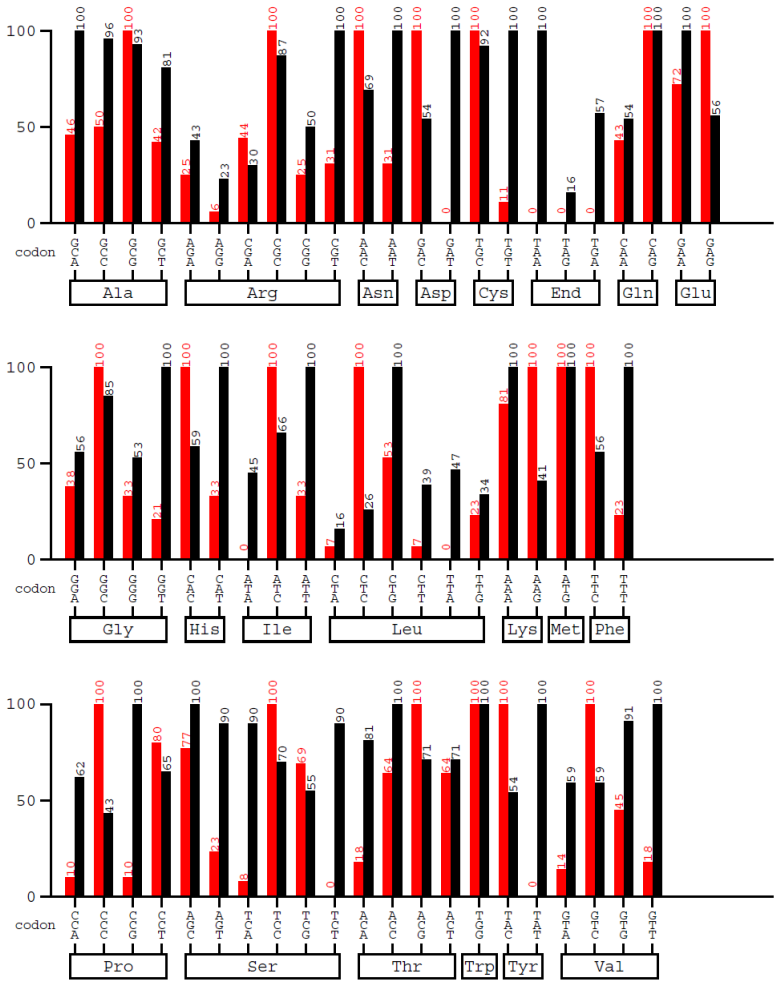


Figure 4.28: Codon usage table created using the gcua software.[196] This shows the relative adaptiveness on the y-axis of the codons shown on the x-axis; the red bars represent the relative usage within the *Tg*SPT gene and the black ones represent the relative usage within *E. coli*. The mean difference is 41.31%.

Therefore Rosetta *E. coli* cells, promote a higher and more successful translation rate of the proline, leucine, isoleucine, arginine and glycine codons. If the cells also contained higher levels of tRNAs for leucine (CTC) and lysine (AAG), expression levels could be further increased.

It is interesting to note that the constructs prefer different expression times. SPT Δ143 expresses at higher levels when the expression cultures are left at 15 °C degrees for 48 hours after the OD_{600} reaches 0.5 whereas SPT Δ158, SPT Δ176 and SPT Δ180 express higher levels of protein when left at 25 °C for 23 - 24 hours after reaching an OD_{600} of 0.5. The expression levels achieved this way are large but could potentially be increased by using Arctic Express^{TM} cells.[132] They grow at

lower temperature leading to slower expression and higher expression levels of more active protein. This could improve yield, as increasing the expression times whilst lowering the temperature proved to have a desirable effect on expression levels.

4.3.1.2 *Tg*SPT - a homodimer

In the eukaryotes studied so far, the first enzyme of the sphingolipid biosynthesis pathway, SPT, is composed of a heterodimer with two closely related subunits. In constrast, there is strong evidence that the *T. gondii* SPT, the target of this investigation, forms a homodimer, more closely resembling its bacterial homologues. The homology model presented in section 4.2.2 based on the structure of the bacterial *Sm*SPT homologue is perfectly consistent with the notion of dimer formation. Furthermore, evidence for homodimer formation is clearly seen in the native gel (Figure 4.16) which shows that Δ158 forms a dimer whereas Δ180 shows a band at the correct mass for a monomer. This may be due to this truncation not being as stable in solution as the others. As the Δ180 is the shortest construct, this suggests that the N-terminus is involved in dimer stabilisation. This is consistent with the observation that the shorter constructs are more prone to precipitation, particularly during attempts to concentrate soluble protein samples.
SPT Δ158 was submitted for mass spectrometry analysis (section 4.2.5). The LC-ESI-MS mass spectrum (Figure 4.17) shows an $[M+H]^+$ peak at 46132 ±1 *m/z* and a $2\times[M+H]^+$ at 92264 ±1 *m/z*, consistent with the calculated mass of SPT Δ158 containing two additional residues (glycine and proline) which remain after tag cleavage. It also provides evidence that the PLP cofactor is no longer bound after ESI, presumably due to it not being covalently bound and therefore not being able to withstand the forces of ionisation. Finally, Bio-SAXS experiments on samples of all four SPT constructs have confirmed homodimer formation in solution. For SPT Δ158 and SPT Δ176, the data correlates well with the expected molecular weight for a homodimer. For the longest, SPT Δ143. and shortest, SPT Δ180, constructs, the data suggest that the former forms aggregates in solution whereas the latter may be transient and exist both as a monomer and a homodimer in solution. However, it could be argued that the data is within the experimental accuracy expected, suggesting that all four SPT constructs exist as homodimers in solution.

In summary, there is strong evidence that the *T. gondii* SPT forms a homodimer; it is evolutionary more closely related to the bacterial orthologues than the eukaryotic orthologues. This is important as the specific differences between eukaryotic SPTs and this enzyme have to be exploited in order to design selective drugs.

4.3.2 *Tg*SPT secondary structure determination by CD spectroscopy

Analysis of the CD spectroscopy data revealed that the the three constructs; SPT Δ158, SPT Δ176 and SPT Δ180, consist of between 33 and 55% α-helices; this value gets smaller with construct size. SPT Δ176, has a smaller α-helix content than the other constructs (at 33%), despite it being the second shortest construct. This may be due to protein sample degradation or aggregation. Even though the values calculated for secondary structure components are mostly useful as a guideline, the trend for the data suggests that SPT Δ176 should contain approximately 54% α-helices. The percentages calculated for the other structural elements of SPT Δ176, when compared to the values obtained for SPT Δ158 and SPT Δ180 suggest that the SPT Δ176 sample quality had been affected.

The secondary structure elements were calculated by three programmes; CONTINLL, CDSSTR and SELCON3, which utilise the reference set input to calculate values of molecular circular dichroism based on the data input into them. Data shown (for percentage of secondary structure elements) was calculated by CONTINLL for the three proteins; this was chosen by comparing the experimentally measured $\Delta\ \epsilon$ with the $\Delta\ \epsilon$ calculated by each of the programmes. The calculated data showing the best overlap with the experimentally measured values was used to predict the percentage of secondary structure elements.

Even though this data only gives an approximation of the structure of the proteins, it can be compared to other SPTs from the PDB; *Sm*SPT contains approximately 45% α-helices and 13% β-sheets. Despite only studying *Tg*SPT truncations, the measured structural elements occur with the same frequency as for this SPT homologue, suggesting that the proteins are properly folded as they have the expected secondary structure percentages.

4.3.3 SPT activity assay

4.3.3.1 Identification of active domain of *Tg*SPT

As SPT catalyses the condensation reaction of palmitoyl-CoA and L-serine, the rate-limiting step of sphingolipid synthesis, an appropriate assay for identifying

SPT activity is to monitor the inclusion of [C^{14}]-serine into KDS, the product of the condensation reaction.
The TLC plate displaying the extracted product shows that the SPT activity of the four constructs decreases with size. The bands for radiolabelled KDS get fainter and are only just visible for the two shortest constructs, SPT Δ176 and SPT Δ180. This suggests that the deletions may be either disrupting protein folding, affecting substrate interaction or formation of the homodimeric complex. Sequence alignment of *Tg*SPT and *Sm*SPT indicates that the longer constructs completely contain the relatively conserved domain whereas the two shorter constructs are lacking part of this domain, potentially contributing to the diminished enzymatic activity. The amino acid sequences alignment with the bacterial SPTs revealed that the N-terminal extension is not present; this domain contains a transmembrane region which does not appear to have any effect on the enzymatic functionality of *Tg*SPT *in vitro*. These results have identified the active domain of *Tg*SPT .
The size of the spots seen for the reaction product, KDS, in Figure 4.21, suggest that SPT Δ143 is less active than SPT Δ158, in contrast with the results seen in Figure 4.23. This may be a coincidence, but could also be linked to the inaccuracies linked to the experimental protocol - in particular the step for product extraction. In order to quantify the amount of product produced by the various constructs, an alternative experiment could be applied. M. F. Rütti *et al*[197] have developed a method where KDS is reduced by $NaBH_4$ and forms erythro- and threo-sphinganine (the reaction is not stereospecific). Ortho-phthalaldehyde is used to derivatize the reduction products which are separated by HPLC and subsequently quantified using fluorescence detection (all based on method developed by Riley *et al.*[198] This method both avoids the use of radio-labelled compounds and optimises the amount of substrate needed.
Employing the standard [C^{14}]-serine SPT activity assay, it is clear that the enzymatic activity decreases with decreasing construct size, but that activity is detectable for all four *Tg*SPT constructs.

Combined with the data obtained from CD spectroscopy and Bio-SAXS, the SPT constructs appear to have the expected secondary structure percentages and form homodimers in solution. The activity assay thus also shows that *Tg*SPT is functionally active as a homodimer, as its prokaryotic SPT homologues and not as a heterodimer as other eukaryotic SPTs (P. W. Denny, personal communication).

4.3.3.2 Inhibition of SPT activity with myriocin

Myriocin (see Figure 4.8) is a known inhibitor of SPT[183] at picomole concentrations. Myriocin (from *Mycelia sterilia)* is an analogue of sphingosine, which possesses antifungal antibiotic and immunosuppressive properties, and it resembles both substrate and intermediate products of the SPT reaction. As such small quantities are required for SPT inhibition, this was chosen for inhibition studies. Other inhibitors (section 4.1.8.1) could be tested in future experiments.

It is clear from the experimental results (section 4.23) that myriocin inhibits the SPT activity of *Tg*SPT constructs SPT Δ143 and SPT Δ158. Complete inhibition of SPT Δ143 is seen and almost complete inhibition of SPT Δ158. The reason for the incomplete inhibition of SPT Δ158 may be due to less effective binding. This could also be a result of the experimental protocol and the relative insolubility of myriocin in CH_3OH.

Calculating the IC_{50} value of myriocin on SPT Δ158 under the experimental conditions employed was not possible; the results obtained were not sufficient to calculate any definite values as the standard deviation error calculated based on the four independent measurements were too high (ranging from 32 - 83%).
Myriocin is only soluble in CH_3OH to a concentration of 2 mg/mL but in DMSO it is soluble to a concentration of 25 mg/mL. The assay employed to determine the IC_{50} value is not optimal, as a serial dilution of the myriocin is required and due to the low solubility, large errors are difficult to avoid. An alternative solvent to use is therefore DMSO.[184]
Myriocin inhibits SPT activity for SPT Δ143 fully and partly inhibits SPTΔ158 enzymatic activity.

4.3.4 Discussion of crystallisation of SPT

No diffracting crystals were obtained for any of the four SPT constructs. The crystals shown in Figure 4.27 were never tested but their decomposition suggests that they may hay have been protein crystals. Several experimental adjustments can be made to improve chances of crystallisation in the future.
Firstly, adjustments to the purifications conditions could be implemented to ensure more homogeneous protein samples. SPT Δ158 and SPT Δ180 showed large difference in homogeneity on the Malvern DLS (data not shown) with the larger construct being more homogeneous in solution than the shorter construct. This is also evident

from observations of protein precipitation occurring for the shorter constructs after 12 hours at 4 °C. To stabilise the shorter constructs, the buffers used for purification could contain a higher percentage of glycerol e.g. 10%. Another option would be to purify the protein using an anion exchange column as well; this was attempted but did not improve protein purity and resulted in loss of protein instead. To counteract this, further stabilising compounds could be added to the purification buffer.

To aid crystallisation of a protein, co-crystallisation with ligands is often successful. One option is to use identified inhibitors or sphingolipid derivatives or even different amino acids such as L-serine or D-serine. Other ligands can be identified using thermal stability assays. Several salts have been identified as having a stabilising effect on Δ158. $Na_2C_3H_2O_4$ is an obvious choice, however a screen was set up with 1 M $Na_2C_3H_2O_4$ added to the protein buffer but the protein precipitated, suggesting that this compound is not so useful for stabilising *Tg*SPT constructs for crystallisation. Na_2SO_4 and $Na_3C_6H_5O_7$ were shown to stabilise SPT Δ158 with the TSA and the ability to increase the ionic strength, especially of $Na_3C_6H_5O_7$, without reaching too high salt concentrations, makes these two viable candidates for stabilising protein during crystallisation trials. Even though $Na_3C_6H_5O_7$ shows less protein-ion interaction than $Na_2C_3H_2O_4$ and shows less increase in stability with ionic strength, it may still be a better option than $Na_2C_3H_2O_4$, as it is easy to increase ionic strength whilst keeping the concentration within an achievable range. Other strategies involve addition of detergents or peptergents using available screens. Other useful screens include solubility screens (in the case where no better buffer system is identified for the protein) and pH precipitants. The Index HT screen (Hampton Research) is a specific example of a screen which is a sparse matrix, grid screen involving factorial screening where many know reagents, novel and traditional, are combined by reagent class. To improve the quality of potential crystals, cryoprotectants can be added to the crystallisation screens.

In some cases addition of trypsin to the soluble protein before setting up crystallisation screens has proved successful, as the protein may cleave in a favourable manner for crystallisation.

Other crystallisation methods such as microbatch under oil, microdialysis or free interface diffusion methods are other options that could be explored but perhaps will not have as large an effect on the formation of *Tg*SPT crystals.

Mutation of the SPT constructs could also be explored as well as other homologues such as *Eimeria tenella* SPT. However, before resorting to those options, seeding or microseeding using the crystals grown for SPT Δ158 could be used and if that fails, then seeding with crystals of homologue crystals can present an alternative to achieving protein crystallisation.

There are many possibilities that have not yet been explored but which could be

promising with stable protein expression and purification protocols in place.

4.4 Chapter conclusions

The work demonstrated in this chapter shows a biophysical and biochemical analysis of the *T. gondii* SPT drawing on multiple techniques.

Homology modelling of *Tg*SPT based on the structure of *Sm*SPT gave rise to a predicted structure of *Tg*SPT where the N-terminal end was missing, but where the conserved domain and active sites were modelled. The resulting model showed a PLP binding site very similar to that of both *Sm*SPT and *Sp*SPT; the corresponding residues had the potential for hydrogen bonding to PLP and they were in the correct conformation, leading to the conclusion that the active site of *Tg*SPT is highly likely to be similar to that of other SPTs and, by inference, so is the reaction mechanism.

Successful transformation of four *Tg*SPT constructs and expression of them using an autoinduction method was demonstrated, resulting in expression levels between 3 and 32 mg of protein per litre of expression medium.

The resulting proteins were verified to be correctly folded by CD spectroscopy. Mass spectrometry experiments on purified SPT Δ158 provided strong evidence for *Tg*SPT existing as a homodimer which was also supported by Bio-SAXS data.

Serine palmitoyltransferase activity was recorded for all four recombinant proteins; the activity levels were seen to decrease with construct size. The product of the condensation reaction occurring between serine and palmitoyl-CoA, KDS, was confirmed by LC-ESI mass spectrometry.

Inhibition with the known SPT inhibitor, myriocin, was also shown for constructs SPT Δ143 and SPT Δ158. The attempt to establish the IC_{50} value of myriocin on SPT Δ158 did not yield a specific result so further experiments are being conducted under different solvent conditions.

Crystallisation trials for the four SPT constructs did not result in any diffracting crystals. However, TSAs identified several stabilising conditions for SPT Δ158 which suggest a starting point for future crystallisation experiments.

Chapter 5

Triosephosphate isomerase from *Sulfolobus solfataricus*

5.1 Introduction

Triosephosphate isomerase (TPI) is an enzyme which catalyses the reversible isomerisation of dihydroxyacetone phosphate (DHAP) to D-glyceraldehyde-3-phosphate (GAP) in the Embden-Meyerhof-Parnas (EMP) pathway of glucogenesis. The enzyme adapts the most common fold in nature, the $(\beta\ \alpha)_8$-barrel, making it the ideal enzyme for studying protein evolution. *Sulfolobus solfataricus* is a hyperthermophilic archaeon, and thus the *Ss*TPI has adapted to the extreme environment in which this organism exists.
The aim of this project was to crystallise the enzyme and solve both an unbound- and a ligand-bound form allowing for comparison with other hyperthermophilic TPI-barrels to further explore the determinants for thermostability.

5.1.1 The archaeon *Sulfolobus solfataricus*

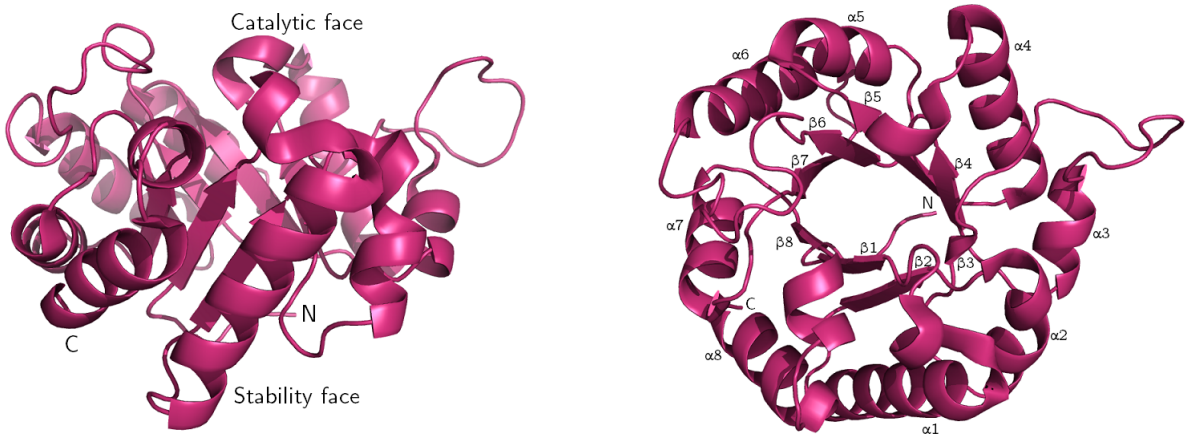
Sulfolobus solfataricus is an extreme thermoacidophilic archaeon (single celled procaryotic microorganism) which belongs to the phylum Crenarchaeota. Many archaea are found living under extreme conditions. *S. solfataricus* grows at temperatures between $70 - 90\,^{\circ}\mathrm{C}$ with optimal growth at $80\,^{\circ}\mathrm{C}$ and at pH 2 - 4 (with an optimal pH of 3.2).[199]
Sulfolobus, the genus, was first isolated by T. D Brock *et al*[200] from acidic hot springs in Yellowstone National Park. Before this discovery, it was believed that

life was not sustainable at extreme temperatures and pH values.[201] Since then, several other species of the genus *Sulfolobus* have been discovered all over the world in sulphur rich soils, marine ecosystems,[202] acidic hot springs and thermal acidic soils.[203] The species *Sulfolbus solfataricus* was isolated by W. Zillig *et al*[199] in 1980 from terrestrial volcanic hot springs near Pisciarelli, Italy. It was originally believed to be a bacterium possessing extraordinary properties, but a new classification system showed that *Sulfolobus* belonged to a new prokaryotic domain, namely the Archaea,[204,205] the only kingdom that can grow at temperatures above 95 °C.[206]
In the *Sulfolobus* species, all strains have been found to grow autotrophically and heterotrophically, making them extremely versatile.[203] *S. solfataricus* has the ability to grow aerobically[203] on a wide range of materials such as peptides, sugars, on plates and in liquid media,[207] which makes it easy to manipulate. *S. solfataricus* is a model organism for Archaea (as the *Sulfolobus* species are so abundant)[203,208] and thus in 2001, sequencing of its genome (2.99 Mbp) was completed by Q. She *et al.*[209] It was chosen for the possibility of gaining an understanding on the molecular basis for thermoacidophilic adaptation, regulation of enzymatic pathways and eukaryotic cellular evolution.

5.1.2 Triosephosphate isomerase structure

Triosephosphate isomerase is one of the most studied and structurally best characterised enzymes. The protein adapts the most common fold for proteins found in nature[210] - the $(\beta\ \alpha)_8$-barrel fold, also known as the TPI-barrel (or TIM-barrel) fold - which was first discovered in TPI.[211] Approximately 10% of all solved protein structures have been found to contain at least one domain with a TPI-barrel fold and due to this prevalence, TPI-barrels make a useful model for studies of function, stability and evolution of proteins.[212]
Most $(\beta\ \alpha)_8$-barrels consist of 200 - 250 amino acids which are divided into eight parts. Each of these subunits are made up of a β-strand connected to an α-helix via a $\beta\alpha$-loop. The motifs are then connected by $\alpha\beta$-loops. A barrel is formed by the eight β-strands that make up a central core which is surrounded by eight α-helices. Each β-strand is slanted with respect to the central barrel by approximately 36 ° which results in a staggered β-sheet,[213] leading to a shift of eight residues for every turn when following the barrel.[210] The approximate 4-fold symmetry of TPI-barrels arises due to the layers formed inside the barrel core by the residue side-chains (usually the side-chains arrange themselves perpendicularly to the barrel axis and form a layer which is three or four thick). Four alternating residues from the odd-

or even-numbered β-strands ((1, 3, 5, 7) or (2, 4, 6, 8)) make up each layer. The stability of the fold is preserved by residues located in the core of the barrel (and at the opposing end to the catalytically active residues) near the $\alpha\beta$-loops (this forms the stability face). The active site residues identified in all the studied $(\beta\ \alpha)_8$-barrels have been located at the C-terminal ends of the β-strands as well as in the $\beta\alpha$-loops (known as the catalytic face), as shown in Figure 5.1.[214,215]



(a) (b)

Figure 5.1: Diagram of *Pyrococcus woesei* TPI monomer.[210] (a) Sideways view with the N- and C-terminals, catalytic and stability faces annotated. (b) View down the barrel with all α-helices and β-sheets numbered.[212]

The structures shown in Figure 5.1 are of the *P. woesei* TPI.[210] This was chosen as the model as it is the closest homologue to the *S. solfataricus* TPI with a sequence identity of 43%.

5.1.3 Enzymatic function of $(\beta\ \alpha)_8$-barrels

Apart from a few exceptions, all studied $(\beta\ \alpha)_8$-barrels are enzymes, functioning either as isomerases, hydrolases, lyases, oxidoreductases or transferases (covering five enzyme classes that the Enzyme Commission (EC) has defined).[212] Approximately two thirds of the $(\beta\ \alpha)_8$-barrels have substrates or cofactors which contain a phosphate group,[216] as the active site contains a phosphate binding site. Furthermore, it has been hypothesised that the catalytic face, and therefore also the active site of the barrel, has a positive potential due to a dipolar electrostatic field associated with the barrel, arising from the α-helices and the microdipoles linked to them. The electrostatic field may be the force behind the $(\beta\ \alpha)_8$-barrel preference for metabolites

with a negative charge.[217]

5.1.3.1 Enzymatic function of TPI

TPI catalyses the reversible isomerisation of the heatlabile intermediates dihydroxyacetone phosphate (DHAP, a ketose) and D-glyceraldehyde-3-phosphate (GAP, an aldose), which is a vital reaction within the glycolysis pathway. The reaction was first characterised by Meyerhof and Kissling,[218] and is described in Equation 5.1.[219,220]

$$E + DHAP \rightleftharpoons E..DHAP \rightleftharpoons E..GAP \rightleftharpoons E + GAP \tag{5.1}$$

where E is the enzyme which associated with DHAP to form GAP after which the enzyme is recovered.
In TPI $(\beta\ \alpha)_8$-barrels, four resides have been identified as being crucial for the isomerisation mechanism catalysed by TPI. The isomerisation reaction is based on acid-base catalysis. The residues relevant for functionality are an asparagine and a lysine in the substrate binding site and a histidine (electrophile) and a glutamate (proton acceptor) in the active site. The asparagine (Asn9) and a lysine (Lys11) residue, both located in the $\beta 1\alpha 1$ loop, facilitate the substrate orientation, whilst a histidine (His93) and glutamate (Glu141) residue, in loops $\beta 4\alpha 4$ and $\beta 6\alpha 6$ respectively, are involved in the proton transfer steps (where the histidine residue acts as the electrophile and the glutamate residue acts as the proton acceptor). A proton is removed from carbon 1 from DHAP by the glutamate. The histidine residue donates a proton to stabilise the negative charge associated with the carbon 2 carbonyl group as a result of the extracted proton. The proton transfer results in the formation of an enediol intermediate. To complete the isomerisation to GAP, a proton is withdrawn from the carbon 1 hydroxyl group by the histidine residue and a proton is added to carbon 2 by the glutamate residue. The isomerisation is thus complete and the enzyme is recovered. The process just described is known as the "classical" mechanism (see Figure 5.2).[212,221,220]

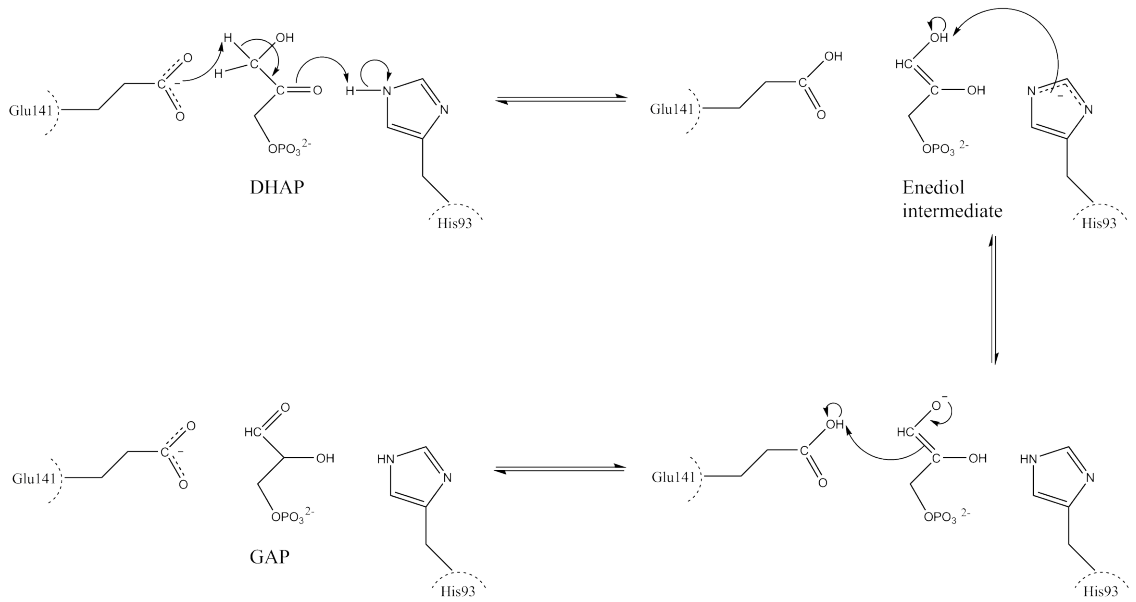


Figure 5.2: Schematic of the "classical" mechanism of TPI. This is shown in the physiological direction which occurs during glycolysis (from left to right). The two residues relevant for the reaction, His93 and Glu141, are shown with the DHAP substrate, the formation of the enediol intermediate, the following proton withdrawal and addition by His93 from carbon 1 hydroxyl group and from Glu141 to carbon 2 respectively. This results in completed isomerisation of DHAP to GAP.

When GAP is isomerised to DHAP, the enzyme behaves perfectly and the rate-limiting step becomes the step where substrate and enzyme encounter each other (which is limited by the rate of diffusion).[221] Simulations have suggested that the GAP is directed by the electrostatic potential of the active site.[212]
Theoretical models of the active site along with NMR experiments[222] have supported the proposition that the glutamate residue replaces the histidine residue's role in the mechanism; the glutamate residue donates a proton to the substrate oxygen atom number 2 forming an enediol intermediate whereafter an enediolate is formed as a proton is donated from the oxygen number 1 of the substrate to the glutamate residue.
An alternative method of isomerisation, also know as the "criss-cross" mechanism, exists for the reaction following initial proton extraction from the carbon 1. This proton is transferred to oxygen 2 on the substrate and not the adjacent substrate carbon. The glutamate is the only residue involved in the catalysis reaction; the resulting negative charge on the substrate is stabilised by the lysine, the asparagine and the histidine that form two oxyanion holes.[220,223] This mechanism was suggested by T. K. Harris *et al*,[224] who observed that both mechanisms play a role for this reaction for the wild-type TPI from *Leishmania mexicana*.[223]
The protein side chain is believed to shift as catalysis occurs, thus suppressing side

reactions. It does not change conformation but simply opens and closes, even when no ligand is bound.[225] In the apo form protein, the $\beta 6\alpha 6$ loop is interacting with the $\beta 5\alpha 5$ loop, but in the holo form, the former loop is positioned on top of the active site, resulting in the glutamate residue re-positioning for catalytic activity as well as allowing for tighter substrate binding.[223]

5.1.3.2 Function of the *S. solfataricus* TPI

S. solfataricus employs a sophisticated, branched central carbohydrate metabolism. For glycolysis (sugar degradation of e.g. galactose and glucose) a branched Entner-Doudoroff (ED) pathway is utilised involving a non-phosphorylative and a semi-phosphorylative branch, whereas for gluconeogenesis, the Embden-Meyerhof-Parnas (EMP) pathway is utilised, in which the *Ss*TPI catalyses the DHAP and GAP interconversion. TPI is of great interest as it controls the flow of DHAP and GAP, the triose 3-phospate intermediates between glycolysis and gluconeogenesis.[226,227]

5.1.4 TPI ligands and inhibitors

A substantial number of molecules have been identified as inhibitors for eukaryotic TPIs, from assays with pea seed TPI (similar experiments have been done involving the *Trypanosoma brucei brucei* TPI).[228] Inhibiting compounds include the intermediate metabolites phosphoenol pyruvate (PEP), 2-P-glycerate, 2,3-P_2-glycerate, 3-P-glycerate and adenosine triphosphate (ATP) which are all intermediates between TPI acting as an isomerase and the production of pyruvate. Fructose-1,6-P_2, fructose-6-P, glucose-6-P, glucose-1-P and 6-P-gluconate (6-phospho-gluconate) are also metabolites in the glycolysis pathway, occurring before TPI is involved, and all bind to and inhibit TPI. As all of these compounds are intermediates in the glycolytic pathway mechanism, it is assumed that they may play a regulatory role and hence inhibit TPI functionality due to the similarity in structure to the substrates. Further metabolites resulting in TPI inhibition are citrate and succinate.[228,229] These compounds have not yet been tested against (hyper)thermophilic TPIs but are expected to have an inhibitive effect on them.

5.1.5 Evolution of the TPI-barrel fold

Enzymes evolve to improve the ability of an organism to thrive when its surroundings change by obtaining different properties (such as new binding or catalytic regions). As 10% of all solved enzyme structures adapt a $(\beta\ \alpha)_8$-barrel fold,[214] which is a very stable fold capable of catalysing multiple reaction mechanisms, the evolutionary relationship of TPIs is of great interest.[212 ,230]

Divergent evolution arises from random mutagenesis reactions and/or gene recombination from a common ancestor, which can be easily replicated in laboratory settings, and convergent evolution involves similar structural changes occurring in different species to obtain a stable fold. Divergent evolution of TPIs is supported by the fact that the structural features are conserved across the TPI-barrels but convergent evolution is supported by the lack of sequence similarity and the variety of functionality.[214] Several studies have been carried out comparing the structure and sequence similarities[231] as well as active and binding sites of multiple hyperthermophilic TPIs.[210,216] Despite the contrast in catalytic properties, the active site of TPI-barrels is always located at the end of the barrel β-sheets towards the C-terminus pointing towards the theory of divergent evolution. Furthermore, the structural conservation of a phosphate binding site in multiple TPIs is a strong indication that all TPI-barrels share a common ancestor and have therefore evolved via divergent evolution.[230,231]

It is impossible to ascertain with absolute certainty which type of evolution is responsible for the development of the huge variety of TPI-barrels, however divergent evolution from a common ancestor has been found to explain most of the data comparisons.[232]

5.1.6 Protein stability under extreme conditions

The TPI enzyme is ubiquitous in nature and its function in the glycolytic pathway is widely conserved across all three domains of life; Bacteria, Eucarya and Archaea. Due to these extreme environmental conditions thermoacidophiles inhabit, the proteins produced within them have higher thermal stabilities than the mesophilic homologues whilst still maintaining protein fold characteristics of specific proteins families.[233,234,235,236] There are several factors contributing to the thermoadaptation mechanism of proteins as discussed below.

5.1.6.1 Structural determinants of thermostability

Protein stability is governed by various factors including hydrogen bonds, van der Waals interactions, salt bridges and the hydrophobic effect.[206] The sum of all these interactions are equal to the conformational stability of the protein which is best described by ΔG, the free energy difference between folded and unfolded states. The free energy difference is usually in the region of 40 kJ mol^{-1}, in comparison to individual residue stabilising interactions which have a magnitude of up to 25 kJ mol^{-1}.[237] Thus, the mutation of a single residue can have a large effect on protein stability, even if no effect is seen on the overall protein structure.[204] Hyperthermophilic proteins may therefore only need a few more stabilising interactions to increase the overall stability. Due to the difficulty of predicting the effect a structural change will have on stability, as the free energy difference is composed both of folded and unfolded states and the fact that not all residues of any protein have been found to contribute to stabilising interactions, the upper limit for conformational stability can not be accurately predicted.
140 °C has been postulated as being the upper limit for protein conformational stability, as hydrophobic interactions are weak above this limit due to their temperature dependence but this only remains an estimate as many other factors contribute to the overall conformational stability.[206] An inverse correlation has been noted between specific activity and conformational stability via molecule flexibility.[204,238] The activity of an enzyme depends on its flexibility and less flexible enzymes tend to be more stable. This is supported by the observations that enzymes from thermophiles are more stable but less active and less flexible than enzymes from mesophiles at any temperature.[206,233,238,239] It is believed that enzymes from mesophiles are more unstable due to their catalytic functionality, where flexibility is essential. An enzyme's stability temperature (temperature above which the protein is no longer stable) is generally not observed to be much higher than that required for functionality, as this would result in less flexibility leading to less functional catalysis. It is a fine balance between achieving optimal functionality as well as reaching high enough stability which is necessary for a useful and long lifetime. The stabilising and destabilising interactions must be balanced in order to ensure enzymatic stability and function. Flexibility affects the stability of an enzyme either regionally or globally. Experiments suggest that stability is affected more by changes in flexibility at certain points in a protein than others.

Even though hyperthermophilic proteins are conformationally stable at high temperatures, there is no single answer as to what influences irreversible degradative processes have on the proteins.[206] Irreversible processes usually occur as a result of

alterations in covalent bonding within the protein. At high temperatures, several processes with high activation energies can occur. These include deamidation of the amine side chain of glutamine and asparagine, oxidation of cysteine, histidine, methionine, tyrosine and tryptophan as well as succinimide formation in aspartic and glutamic acid and all may be involved in protein inactivation.

5.1.6.2 Thermostability in hyperthermophilic TPIs

In order to gain a better understanding of the factors governing the stability of proteins, the structures of several mesophilic and hyperthermophilic enzyme orthologues have been compared to gain a better understanding of the guiding factors of protein stability.[206] Proteins from hyperthermophilic organisms can be stabilised via many routes.
TPI enzymes form dimers in Bacteria, Eukarya and mesophilic Archaea, but tend to form tetramers (usually a dimer of dimers) in hyperthermophilic Archaea. An increased oligomerisation state has therefore been identified as a potential mechanism for hyperthermophilic proteins to fold and function correctly at elevated temperatures.[233] The hyperthermophilic TPI monomers are usually 10% shorter than the bacterial and eukaryal TPI monomers; TPIs from archaea are generally between 215 - 230 residues compared to mesophilic TPIs of 240 - 250 residues.[240,219] Another common known factor is the formation of more salt bridges. Additionally, disulfide bonds, a reduced number of thermolabile residues (such as asparagine and glutamine), more hydrogen bonds between side-chains and maximisation of the overall electrostatic potential all stabilise proteins at higher temperatures.[233]

The studies of the stable TPI-barrel and its interface interactions have been conducted comparing the inter- and intrastructural interactions. The interactions can be divided into inner and outer barrel interactions. Interactions in the inner barrel involve residues that have side-chains pointing towards the barrel core.[240] These residues usually have bulky side-chains such as leucine, isoleucine and aromatic residues, thus holding the barrel β-sheets together with non-polar interactions between the hydrophobic side-chains. All dimeric TPIs have a tyrosine residue pointing towards the inner barrel, but this is replaced by a glycine in archaeal TPIs, but counteracted by the presence of an aromatic tyrosine residue, instead of alanine, glycine or serine residue in dimeric TPIs. It seems that the mutation involving a compensation is necessary for stability in tetrameric TPIs.[240] Outer barrel interactions involve the residues with side-chains pointing out from the barrel interacting with the residues in the helices and loops. No aromatic residues are present in

the outer barrel in tetrameric TPIs, and all the bulky aromatic resides present in dimeric TPIs are replaced by small, hydrophobic residues, resulting in a more densely packed structure. Overall, hydrophobic interactions and a few hydrogen bonds between residues located at the end of strands and main chain residues contribute to stabilising TPIs. The number of inner and outer barrel interactions in archaeal TPIs does not fully explain their thermostability, but the mutations of several key residues, as mentioned above, influence the compactness of TPIs, which may have an impact on protein stability.

Different protein families adapt to high temperatures through different structural solutions. The structural basis of thermostability is not clearly defined, but includes oligomerisation, hydrophobic interactions and hydrogen bonding, a larger number of subunit contacts (which often manifests itself as higher oligomerisation states), polar interactions between surface and accessible surface regions,[241] decreased loop size, decrease in the amount of protein cavities, resulting in tighter packing of the protein.[240,242]
It is impossible to define any one characteristic as the defining factor of protein stability, as several elements contribute to enhancing stability in the different TPIs.

5.1.7 Archaeal TPIs

There are currently structures available for TPIs from three different Archaea available in the RSCB Protein Data Bank (RSCB PDB); these are TPI from *Pyrococcus woesei*[210] (PDS ID 1HG3), *Methanocaldococcus jannaschii*[240] (PDB ID 2H6R) and *Thermoproteus tenax*[219] (PDB ID 1W0M). A Swiss Model Repository[243] search identifed four TPIs with sequence identities greater than 30% to *Ss*TPI. The top four solutions generated were for the TPIs from *P. woesei*, *M. jannaschii*, *T. tenax* and *T. molitor* (PDB ID 2I9E), with sequence identities of 43%, 40%, 40% and 34% respectively. Out of these four results, the first three are from Archaea but the latter, *Tenebrio molitor*,[244] is a yellow mealworm beetle and was thus left out of structural comparisons.
The TPI enzymes from *P. woesei*, *M.jannaschii* and *T. tenax* are all tetramers, albeit with different properties as soluble proteins. The *Tt*TPI enzyme adopts an equilibrium and reversibly dissociates from an active tetramer to two inactive dimers, suggesting that the interface interactions are weaker than for the other archaeabacterial TPIs (it has been found to be less stable than the *Pw*TPI).
Figure 5.3 shows a sequence alignment of the four archaeal hyperthermophilic TPIs; *Pw*TPI, the *Mj*TPI, the *Tt*TPI and the *Ss*TPI. The sequences are all compared to the secondary structure of the *Pw*TPI (displayed above the alignment). The se-

quence alignment highlights the highly conserved residues and the active site residues amongst the compared TPIs.

Several structure-based sequence alignments have been carried out between mesophilic and hyperthermophilic TPIs (dimeric and tetrameric enzymes) from various organisms. Gayathri *et al*[240] highlighted the large insertions at the sites coinciding with the tetramer interface in the dimeric TPIs. Insertions tend to occur in loop 4, near helix 5 and on helix 6. Additionally, both helix 5 and loop 5 are shorter in tetrameric TPIs than in dimeric TPIs, causing the helix to be closer to the barrel, thus making the structure more compact. Structural comparisons have shown that the insertions have a preventative effect on tetramer formation. Another contributing factor is the charge of the surface residues; the region which corresponds to the tetramer interface in archaeal TPIs is negatively charged in dimeric TPIs (containing residues such as glutamic acid and aspartate), resulting in strong repulsive interactions, if two dimers were to attempt to form a tetramer. In tetrameric TPIs, the corresponding regions are mainly hydrophobic, resulting in a modified surface which favours tetramerization.[240] *Pw*TPI and *Mj*TPI both have a higher alanine content in helix 5, suggesting that this residue has a stabilising effect on the helix after truncation.[210] In *Tt*TPI, which exists either as an active tetramer or inactive dimer, this interface is not conserved, showing that all archaeal TPIs do not possess the same stabilising mechanisms. This may be the reason why the latter TPI exists as a weaker TPI.

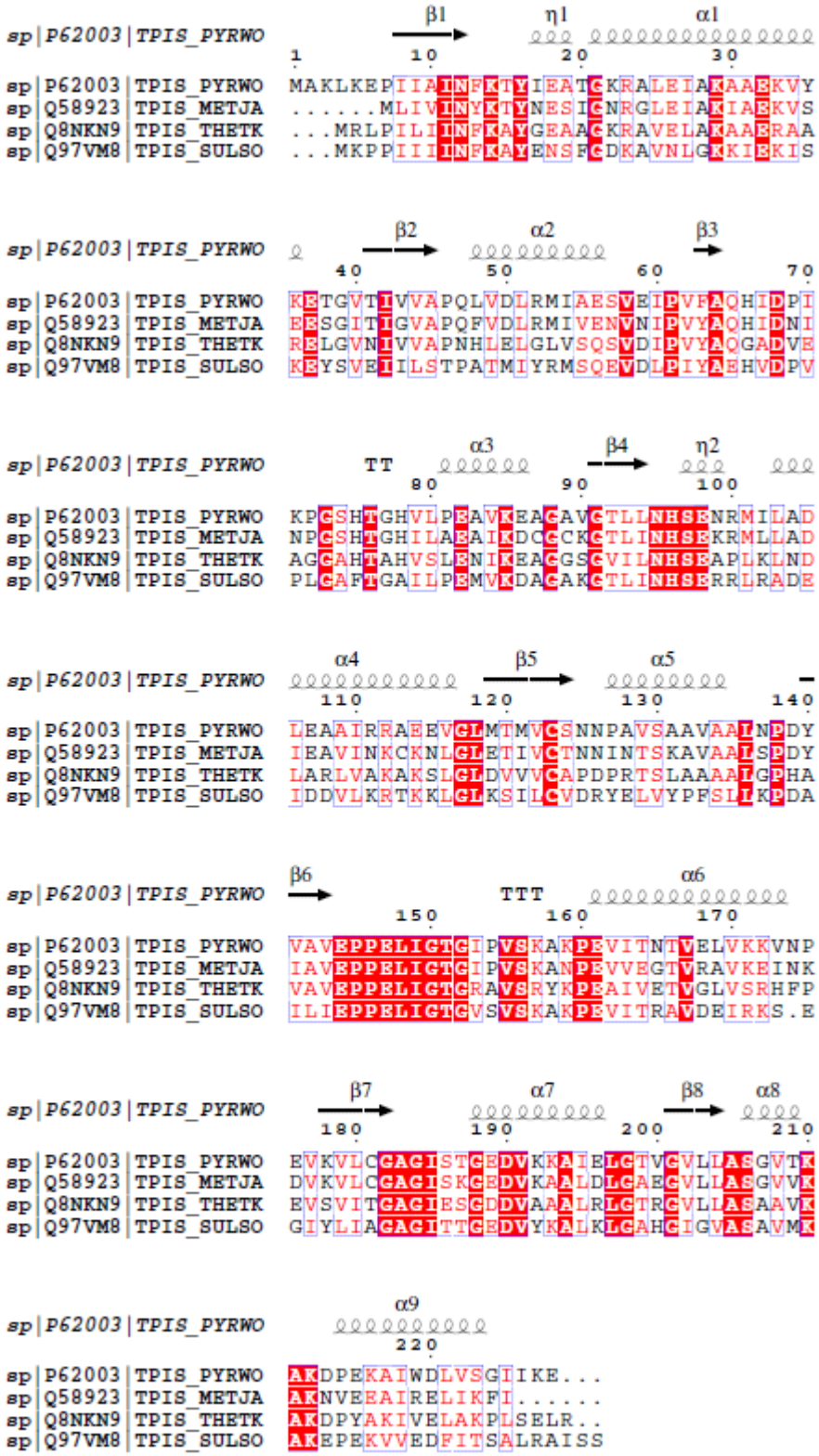


Figure 5.3: Sequence alignment of four thermophilic TPI sequences of *Pw*TPI, *Mj*TPI, *Tt*TPI and *Ss*TPI, produced in ESPript 3[121] (available at http://espript.ibcp.fr). Residues enclosed in a red block are identical, residues enclosed in blue frames with red letters are of similar type and the rest have a low identity score. α denotes α-helices, β denotes β-sheets, TTT denotes strict α-turns and η denotes 3_{10}-helices.

5.1.8 Chapter objectives

In this chapter, the crystallisation of triosephosphate isomerase from the hyperthermophilic archaeon *Solfolobus solfataricus* is presented. *Ss*TPI was crystallised (without and with bound phosphoenol pyruvate (PEP)) and the the 3D structures determined. TPI adapts the most common protein fold, the $(\beta\ \alpha)_8$-barrel, making it a model enzyme to study for evolutionary purposes. The two crystal structures without any ligands bound in the active site reveal that, as in other hyperthermophilic archaea, *Ss*TPI forms a tetramer in solution, in order to stabilise the enzyme under the extreme conditions it is exposed to. Furthermore, the PEP bound form shows conservation of the active site residues when compared with other TPIs from hyperthermophilic organisms.

5.2 Methods

5.2.1 Protein expression and purification

Purified *Ss*TPI was obtained Dominik Esser, and Bettina Siebers (Department of Chemistry, Molecular Enzyme Technology and Biochemistry, at the University of Duisburg-Essen, in Essen, Germany), who had expressed and purified the protein recombinantly in *E. coli*. Briefly, cells were lysed in 0.1 M HEPES/KOH pH 6.5 at 70 °C. Protein was obtained by heat precipitation at 80 °C whereafter it was dialysed against 20 mM HEPES/KOH pH 6.5. The sample was loaded onto an anion exchange-sepharose column and eluted using a 1 M NaCl linear gradient. The relevant fractions were purified further on a HiLoad (26/60) gel filtration column followed by HIC (Butyl sepharose). In the final step, the soluble protein was dialysed against 20 mM HEPES pH 7, 100 mM NaCl.
The purity of all protein samples supplied by D. Esser and B. Siebers was verified by SDS-PAGE as described in section2.5.6.

5.2.2 Protein crystallisation

HTP screening was undertaken using a Screenmaker 96+8^{TM} *Xtal* liquid handler from Innovadyne technologies Inc. to produce sitting drops. Two sitting drops of 100 nL reservoir solution + 100 nL protein solution and of 200 nL reservoir solution + 100 nL protein solution respectively, were set up using the JCSG+,[55] Pact Premier,[55] Structure Screen I and II[59,60] and Clear Strat 1 and 2[58] crystallisation screens. After two weeks crystals were found in two conditions of the sparse matrix JCSG+ screen:[55]
Condition A2: 0.1 M sodium citrate, pH 5.5, 20% PEG 3k.
Condition A5: 0.2 M magnesium formate, 20% PEG 3350.

Optimisation screens were designed around condition A5 (JCSG+), spanning concentrations of 0.1 - 0.2 M magnesium formate whilst keeping the PEG 3350 concentration constant at 20%.
Four optimisation screens (24 conditions each) were set up using the hanging drop trays with a reservoir volume of 1 mL and varied drop sizes. The protein concentration was 6.3 mg/mL. For drops containing phosphoenol pyruvate (PEP), varying amounts of 5 mM PEP in dH_2O was added to the drops. All trays were kept at 19 °C.

5.2.3 Data collection

Synchrotron radiation datasets were collected at the Diamond Light Source (DLS) synchrotron in Oxfordshire, UK (Diamond House, Harwell Oxford, Didcot, Oxfordshire, OX11 0DE) on the macromolecular crystallography beamlines, I02 and I03 (Pilatus 6M-F detectors).[245] Several crystals were tested.
Data were collected from a single crystal for each of the three crystal forms obtained (to resolutions of 2.48 Å 2.28 Å and 2.10 Å respectively). Details of the data collection statistics can be found in Table 5.1 below.

	Crystal form I	Crystal form II	PEP bound
Beamline	DLS I02	DLS I02	DLS I03
Wavelength (Å)	0.9795	0.9795	0.9790
Energy (keV)	12.658	12.658	12.664
Omega start (°)	155.0	220.0	0.0
Oscillation range (°)	0.1	0.1	0.1
Exposure time (s)	0.1	0.2	0.1
Detector distance (mm)	510.7	464.4	433.7
Beam transmission (%)	10	20	20
Beam size at sample horizontal (μm)	100	100	80
Beam size at sample vertical (μm)	25	25	20

Table 5.1: Summary of data collection statistics.

Datasets for crystal forms I and II were collected by M. Grøftehauge and the PEP bound dataset was collected by E. Pohl.

5.2.4 Data processing - structure solution and refinement

DLS datasets were (auto-)indexed in Mosflm[124] using 3 images, thus finding the basis vector and assigning Miller indices. At this stage, the unit cell constants were refined. The Bravais lattice with the highest symmetry and lowest penalty is usually the correct one, however the symmetry can not be determined with absolute certainty at this point in the process.
All diffraction data were integrated using Mosflm.[124] Scala from the CCP4 package[246] was used to scale and merge intensities. Once all reflections have been converted to structure factors, using $C_{truncate}$, R_{free} flags were added. Phaser MR[126] (part of the CCP4 package)[246] was used for maximum likelihood phasing of the data.

A Swiss Model Repository[243] search identified the TPI from Pyrococcus woesei[210] (*Pw*TPI) (PDB code 1HG3) as having a 43% identity to *Ss*TPI. The pdb file for *Pw*TPI was trimmed to only contain one chain (ie. one monomer of the *Pw*TPI tetramer) in Pymol),[192] in order to search for a single asymmetric unit (asu). For datasets 2 (crystal form II) and 3 (PEP bound), a monomer of the structure of dataset 1 (crystal form I) was used as the search model. Phenix[247] (1.8.1-1168) was used to refine the models built in Coot[191] (version 0.7). This was done using iterative cycles of model building in Coot followed by refinement using phenix.refine. During the refinement process, the Ramachandran and rotamer outliers were carefully monitored. Water molecules were assigned manually for peaks larger than 3σ with appropriate geometry.

All crystallographic parameters derived from the structure solution process can be found in Table 5.3 in section 5.3.3.

5.3 Results and Discussion

5.3.1 SDS-PAGE analysis of *Ss*TPI

S. solfataricus TPI protein was received at a concentration of 6.2 mg/mL in a 20 mM HEPES pH 7 and 100 mM NaCl buffer. Protein purity was assessed by SDS-PAGE (section 2.5.6) as shown in Figure 5.4.

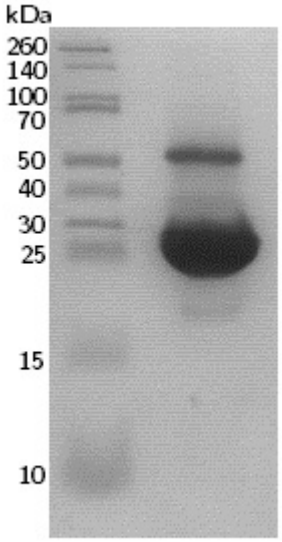


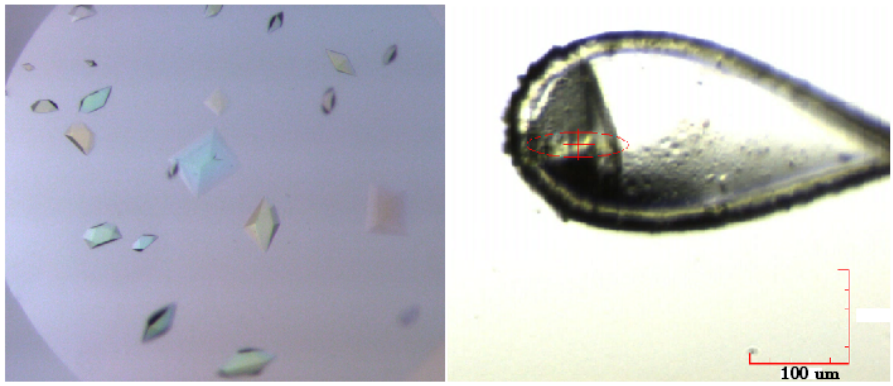
Figure 5.4: SDS-PAGE showing purified *Ss*TPI. 31μg of purified protein resolved on a 12% SDS-PAGE gel. Proteins were stained with Instant Blue (Expedeon) and protein size was estimated relative to Thermo Scientific^{TM} Spectra Multicolour Broad Range Protein ladder.

Figure 5.4 shows approximately 95% pure *Ss*TPI resolved on a 12% SDS-PAGE gel. The lower, broader band at approximately 25 kDa is *Ss*TPI and the higher band at approximately 50 kDa represents dimerised protein (*Ss* TPI consists of 227 amino acids and has a molecular weight of 24736 Da.)

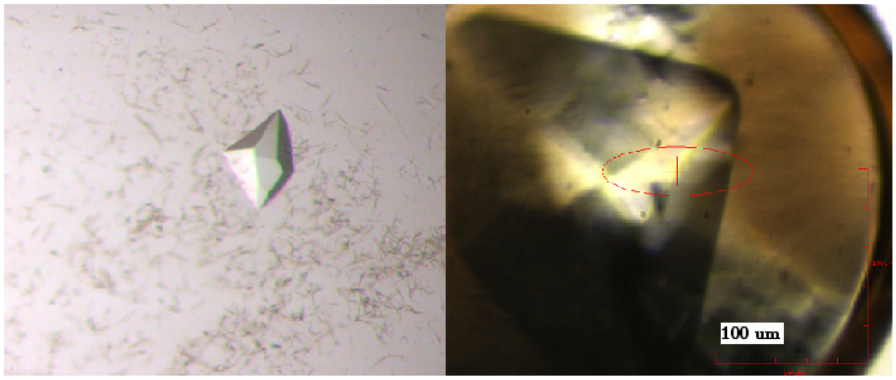
5.3.2 Crystallisation conditions for *Ss*TPI

Initial crystallisation conditions were were identified in condition A2 and A5 in the JCSG+[55] crystallisation screen. After collecting a full data set for crystals from condition A5, optimisation screens focussed on similar conditions. The diffracting crystals grew in various shapes. Sizes ranged from 100 $\times$ 150 μm to 200 $\times$ 300

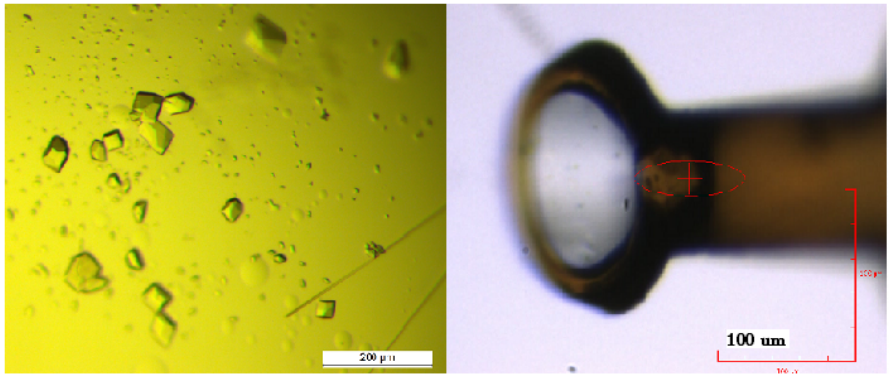
μm for unbound crystals forms, whereas the PEP bound crystals were smaller at 50 $\times$ 75 μm. The relevant crystals are shown in Figure 5.5 for the three conditions yielding diffracting crystals and resulted in full data sets.



(a)



(b)



(c)

Figure 5.5: Pictures of *Ss*TPI crystals in the crystallisation drops and the mounting loops respectively. a) crystal form I; b) crystal form II and c) PEP bound.

The corresponding conditions which produced diffracting crystals and full data sets are summarised in Table 5.2.

Crystal structure	Crystal form I	Crystal form II	PEP bound
*Ss*TPI concentration (mg/mL)	5.9	6.2	6.2
Magnesium formate (MgCHOO) (M)	0.2	0.18	0.16
PEG 3350 (%)	20	20	20
Drop ratio (protein : PEP : reservoir (μL))	0.1 : 0 : 0.1	1 : 0 : 1	1 : 2 : 1
Cryo-protectant (glycerol (%))	-	1.25	10

Table 5.2: Summary of the crystallisation conditions for the diffracting *Ss*TPI crystals.

5.3.3 Structure determination for *Ss*TPI

Data were collected and refined for three different structures. Two of these were crystal modifications of unbound TPI and one was a holo structure of TPI bound to PEP.

The X-ray data and refinement statistics are summarised in Table 5.3. The statistics calculated by RAMPAGE[248] (available at http://www-cryst.bioc.cam.ac.uk/rampage/, utilised to generate Ramachandran plots), concerning the position of the individual residues are summarised at the bottom of Table 5.3.

All structures are well-refined as shown by the R-factors ranging from 16.1 to 21.5%. The bond length rmsd (bond) values are between 0.007 and 0.009 Å with excellent geometry.[249] For crystal form I, the Ramachandran plot (Figure 5.6) reveals a proline outlier (Pro157, chain A). It is the only residue falling into the outlier region. The resolution of the structure is not high enough to determine its exact conformation. Six residues are located in the allowed region, corresponding to 2.8% of the residues. This is higher than the ideal 2% but falls within the experimental error. Pro157 is located near the poorly modelled region where residues 148 to 150 are missing, and near the active site residue number 141.

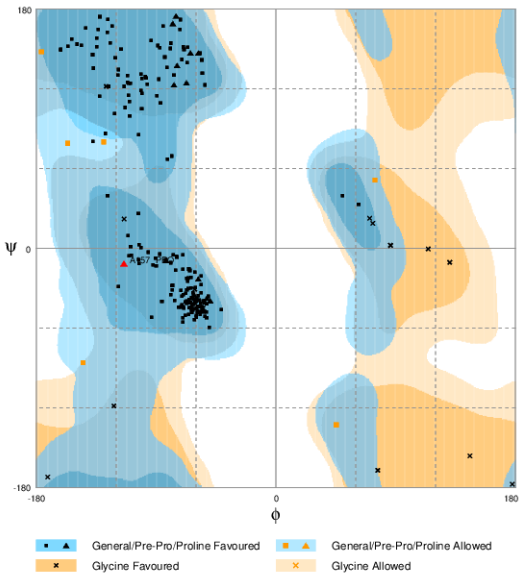


Figure 5.6: Ramachandran plot generated in RAMPAGE[248] of the *Ss*TPI crystal form I.

For crystal form II, the Ramachandran plot (Figure 5.7) shows one alanine outlier (Ala150, chain B).

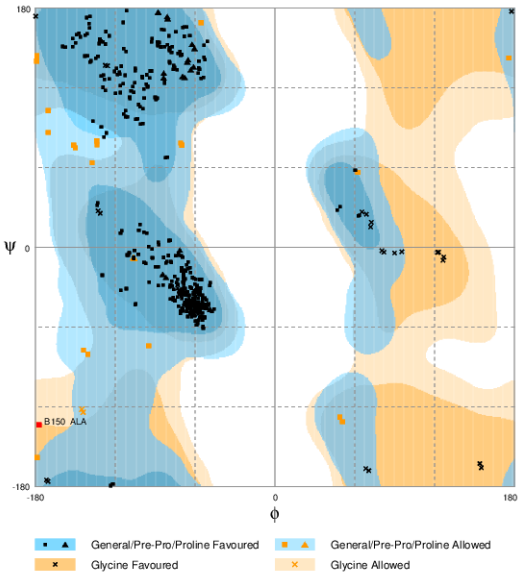


Figure 5.7: Ramachandran plot generated in RAMPAGE[248] of the *Ss*TPI crystal form II.

The Ramachandran plot for the PEP bound structure (Figure 5.8) shows two outlier alanine residues (Ala179, chains A and B). This residue is facing the PEP

binding site in both chains and is pushed away to a distance of approximately 3 Å by the double bonded carbon due to unfavourable interactions making this a structural feature. Only 10 residues are located in the allowed regions, corresponding to 2.2% of the residues.

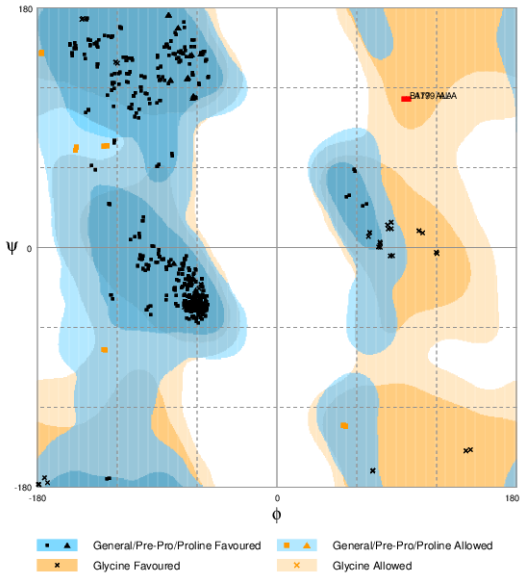


Figure 5.8: Ramachandran plot generated in RAMPAGE[248] of the PEP bound *Ss*TPI.

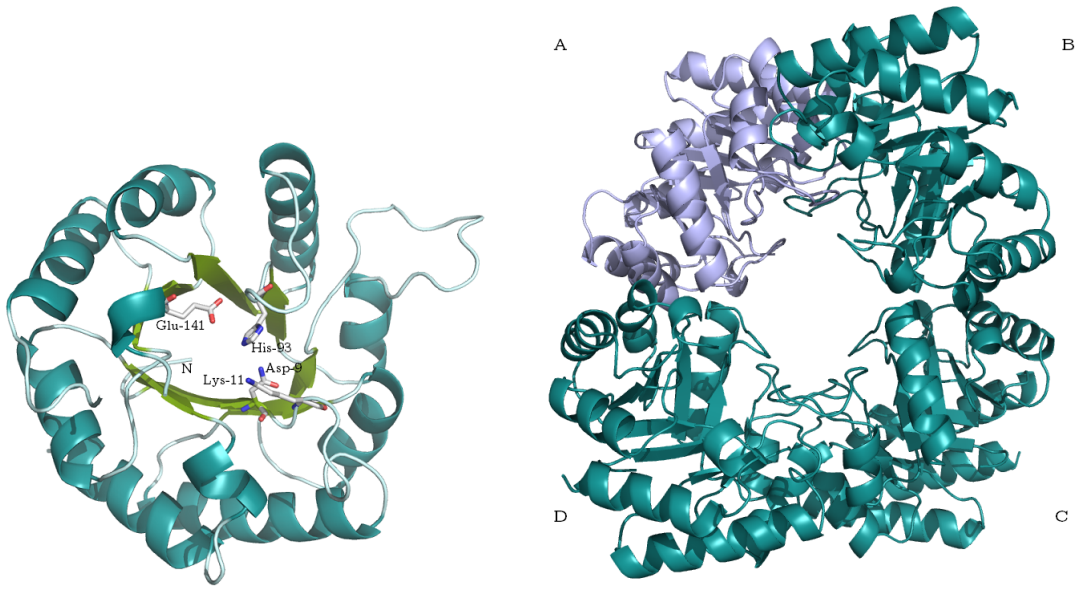
	Crystal form I	Crystal form II	PEP bound
Lattice	Orthorhombic	Orthorhombic	Orthorhombic
Space group	I_{222}	$P_{2_12_12}$	$P_{2_12_12}$
a (Å)	57.1	100.6	57.01
b (Å)	91.6	60.2	90.95
c (Å)	102.2	91.2	99.45
α, β, γ (°)	$90, 90, 90$	$90, 90, 90$	$90, 90, 90$
Resolution range (Å)	51.08-2.48	51.67-2.28	49.51-2.10
Reflections	9722	25614	30328
R_{merge}(%)(∗)	5.6 (17.7)	13.5 (63.6)	13.6 (77.0)
I/σ(I)(∗)	15.4 (7.1)	7.3 (2.6)	6.7 (1.7)
Completeness (%)	98.97	98.82	98.37
Wilson B factor (Å^2)	27.84	36.49	30.15
R (workingset+test set)(%)	16.4	21.9	19.1
R (working set)(%)	16.1	21.5	18.9
R_{free} (%)(based on 5% of data)	22.4	29.2	23.7
No of residues	1-147, 151-223	A: 1-146, 149-226 B: 1-146, 149-226	A: 1-226 B: 1-225
No of PEP	0	0	2
No of waters	151	215	286
Rmsd bond length (Å)	0.007	0.009	0.008
Rmsd bond angles (°)	1.103	1.226	1.121
Number of residues in favoured region	209 96.8%	416 94.3%	435 97.3%
Number of residues in allowed region	6 2.8%	24 5.4%	10 2.2%
Number of residues in outlier region	1 0.5%	1 0.2%	2 0.4%
Outlier residue(s)	Pro157	Ala150 (B)	Ala179 (A and B)

Table 5.3: Diffraction data and refinement statistics for the *Ss*TPI crystal form I, crystal form II and PEP bound crystals. *The number in parentheses refers to the highest resolution shell. The bottom of the table contains Ramachandran plot statistics generated in RAMPAGE[248] stating the number of residues in each region of the plot for the three *Ss*TPI crystal structures. The residues in outlier regions are also specified.

5.3.4 Overall *Ss*TPI structure

5.3.4.1 Crystal form I

The overall structure of *Ss*TPI in the three crystal forms is essentially similar, and therefore the structure will be described based on the *Ss*TPI crystal form I. The crystal form I structure of *Ss*TPI was refined at a resolution of 2.48 Å in the space group I_{222}. There was no clear electron density for residues 148 - 150, so they were not modelled. Residues 224 to 227 (C-terminus) were not modelled, either. One monomer was found in the asymmetric unit as shown by the 3-dimensional representation in Figure 5.9 (a). The structure of it forms a tetramer (arranged as a dimer of dimers) (Figure 5.9 (b)). The *Ss*TPI crystal form I shows the expected, classical TPI-barrel fold, as shown in Figure 5.9 below. The active site resides at the catalytic face of the $(\beta\ \alpha)_8$-barrel, near the C-terminus, are shown and are all in the expected conformation pointing towards the centre of the barrel (Figure 5.9 (a)).

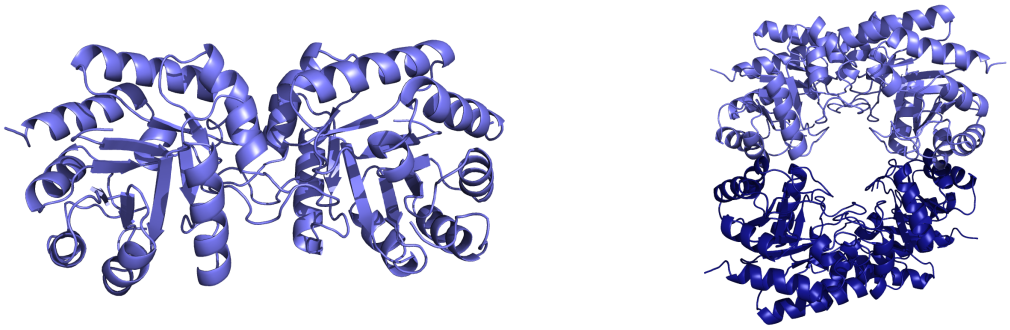


(a) (b)

Figure 5.9: (a) View down the barrel of the *Ss*TPI crystal form I with the four active site residues highlighted. (b) Tetramer of the *Ss*TPI crystal form I. Figures were prepared using *PyMOL*.[192]

5.3.4.2 Crystal form II

The crystal form II structure of *Ss*TPI was refined at a resolution of 2.28 Å in the space group $P_{2_12_12}$. A few more residues were built into this model, and only residues 147, 148 and 227 are missing from the model in both chains A and B. A dimer was found in the asymmetric unit (Figure 5.10 (a)). Figure 5.10 (b) shows the dimer of dimers. Apart from the region surrounding residues 148 to 150 as well as residues 224 to 226, the overall fold is identical to that described for crystal form I (section 5.3.4.1).

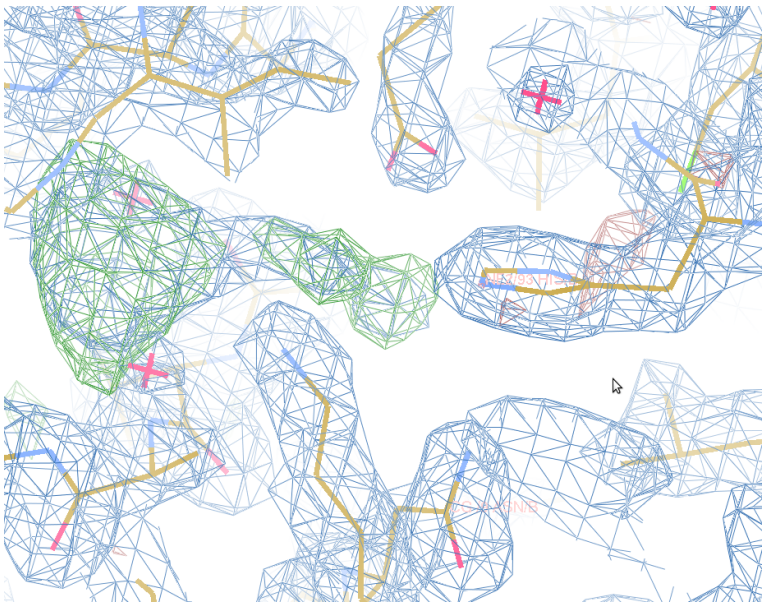


(a) (b)

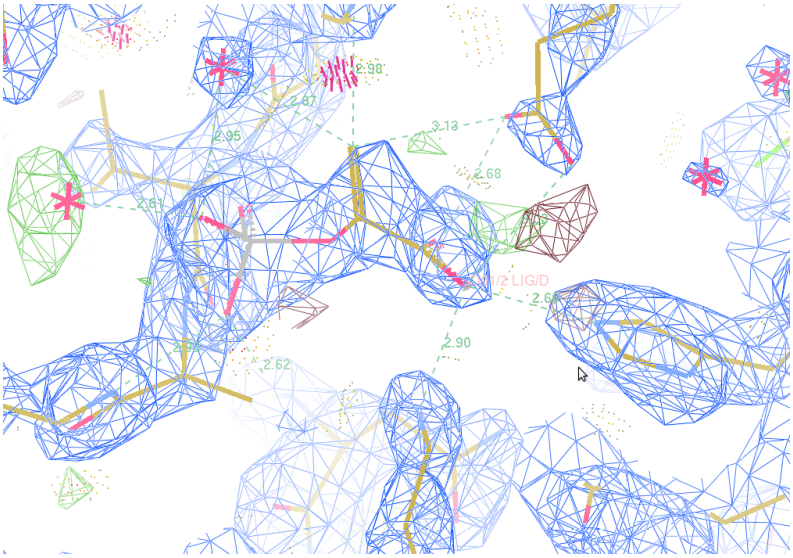
Figure 5.10: (a) Sideview of the *Ss*TPI crystal form II dimer. (b) Tetramer of the *Ss*TPI crystal form II. Figures were prepared using *PyMOL*.[192]

5.3.4.3 PEP bound structure

The PEP bound *Ss*TPI was refined at a resolution of 2.10 Å in the space group $P_{2_12_12}$. The structure is almost fully modelled only missing residue 227 from chain A and residues 226 and 227 from chain B. A dimer was found in the asymmetric unit (Figure 5.10 (a)). The dimer of dimers is shown in Figure 5.10 (b)). Apart from loop 6, which is fully modelled, due to a PEP molecule being bound in the active site, thus closing the loop and creating stability, the overall fold is identical to that described for crystal form I (see 5.3.4.1).
The unbiased $2F_o$ - F_c electron density before the ligand was added is shown in Figure 5.11 (a) and the density with the final model added (stick representation) is shown in Figure 5.11 (b). It clearly shows a large green density where the PEP ligand fits (shown in Figure 5.12).



(a)



(b)

Figure 5.11: (a) The unbiased $2F_o$ - F_c electron density at 1σ level before the ligand was added. (b) The $2F_o$ - F_c electron density at 1σ level with the final model in stick representation superimposed in the active site density. H-bonds are also indicated. Figures are screenshots from Coot.[191]

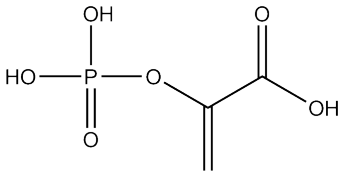


Figure 5.12: Structure of phosphoenolpyruvic acid (PEP) with phosphoenolpyruvate as the anion.

The binding sites of PEP in the tetramer are shown below in Figure 5.13.

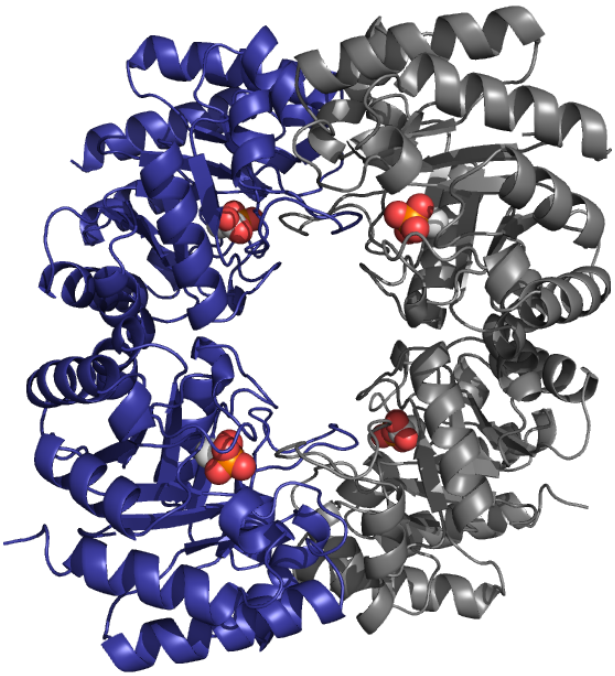


Figure 5.13: The *Ss*TPI tetramer with PEP bound in the active site of each monomer. Figure was prepared using *PyMOL*.[192]

A close-up of the active site in Figure 5.14 reveals the intermolecular interaction between the ligand and the active site residues.

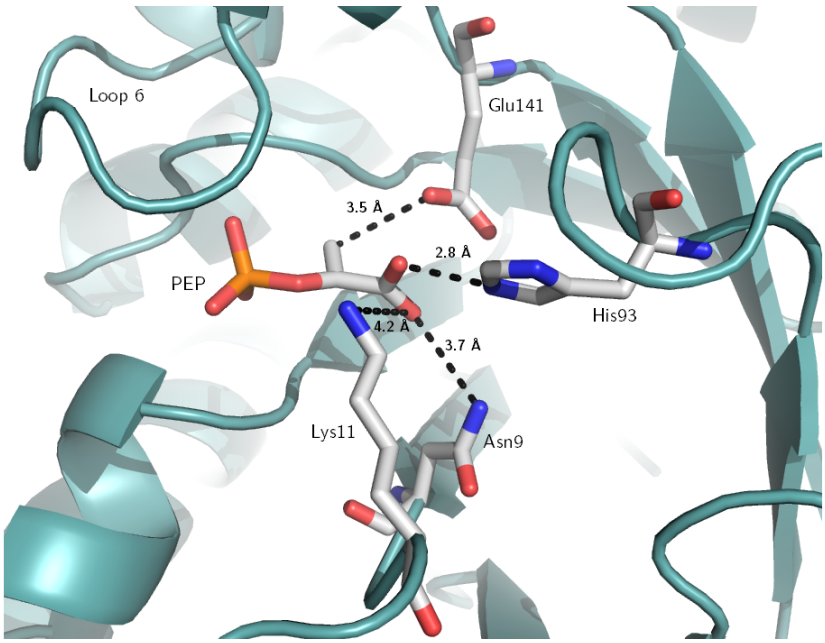


Figure 5.14: Close-up of active site showing the binding site and active site residues, the PEP molecule as well as interaction distances. Figure was prepared using *PyMOL*.[192]

The binding site consists of Asn9 and Lys11 where the interactions occur with the oxygen in the -COOH group. The Lys11 residue could also potentially interact with the O atom of the phosphate group which links it to the rest of the PEP molecule. Loop 6 has been fully modelled and is shown to have closed in, thus keeping the ligand in the active site. It is important to note that PEP is structurally very similar to the substrate, DHAP, and can thus interact with the active site in a similar manner.[250] The active site residues, His93 and Glu141, are also positioned close enough to PEP for interactions occur.
The binding site is largely conserved compared to native rabbit muscle TPI,[250] the only solved PEP-bound TPI structure. A direct comparison, however, is impossible due to the size difference between mesophilic and hyperthermophilic TPIs, but the conserved nature of the binding site suggest evolution of TPIs from a common ancestor.

5.3.5 Structural comparisons of the *Ss*TPI crystal forms

A superposition of Crystal I form, Crystal II form and the PEP bound form of *Ss*TPI shows that the structures are nearly identical apart from the orientation of the active site residue Glu141, and loop $\beta 6\alpha 6$, as shown in Figure 5.15.

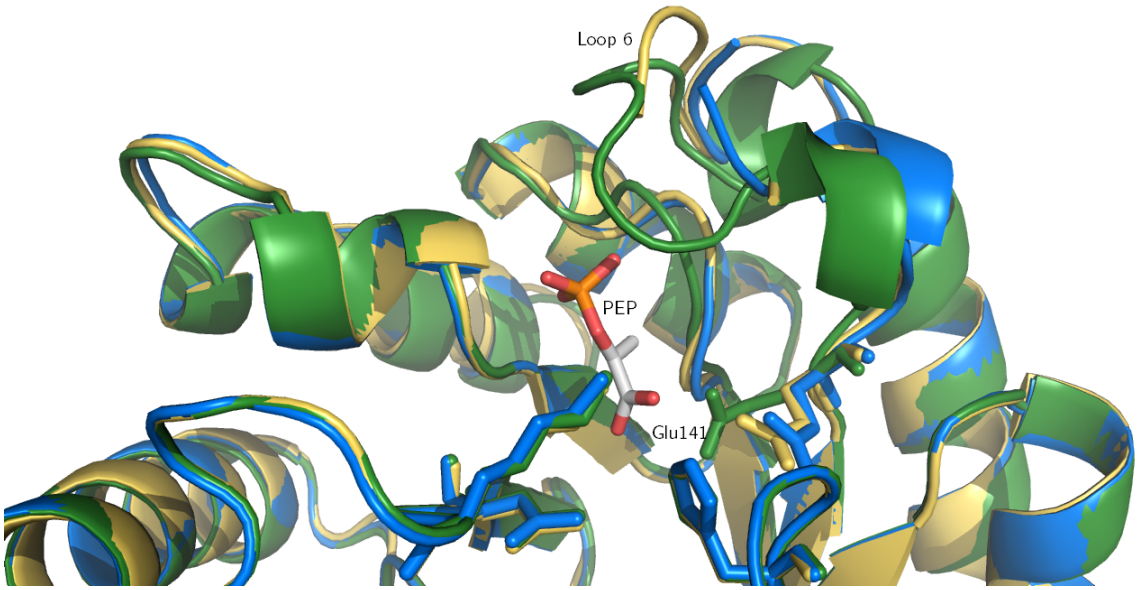


Figure 5.15: Superposition of all three *Ss*TPI structures with crystal form I (blue), crystal form II (yellow) and the PEP bound structure (green). A PEP molecule is shown in the active site and the active site and binding site residues are also shown. The Figure was prepared using *PyMOL*.[192]

The B-factors for the main chain atoms of crystal form I for residues 147 to 151 is 50.0 to 61.7 Å^2. For the main chain atoms of crystal form II for residues 146 to 151 chain A, the B-factors range from 84.5 to 146.2 Å^2 and for chain B from 99.9 to 151.3 Å^2. The B-factors for the PEP bound structure main chain atoms 147 to 151 range between 21.4 and 32.1 Å^2 for chain A and between 31.3 to 39.4 Å^2 for chain B. The values are significantly higher than the overall average in crystal form II for the region between residues 147 and 151 than for crystal form I and the PEP bound form, indicating significant disorder which could be due to the different environment in the crystal structure or to an inherent mobility of these resides.
When comparing the two unbound forms and the PEP bound structure of *Ss*TPI, loop $\beta 6\alpha 6$ is of particular interest. Loop $\beta 6\alpha 6$ is located near the active site and is the link between helix 6 and strand 6 in all analysed TPI-barrels. It has been identified in two different conformations - open or closed.[212] The closed form is usually found in TPI structures containing a ligand and the open form for unliganded structures. In the two unbound forms of *Ss*TPI structures, the loop does not appear to change much. For crystal form II, the loop appears to be closer in conformation to that of the PEP bound structure, which may be due to the low data quality for this part of the structure. It could also suggest that this crystal form and space group forces the loop to be more ordered. Only in the PEP bound form has it been fully modelled. This suggests that the loop may be switching between open and

closed conformations in the unbound structures, but is only fully stabilised in the closed conformation for holo structures, in this case the PEP bound structure, by the PEP molecule.[240,251]
This is all due to the side chain shifting as catalysis occurs in order to suppress side reactions. The $\beta 6\alpha 6$ loop and the $\beta 5\alpha 5$ loop interact in the unbound structures, but when a ligand is bound in the active site, the $\beta 6\alpha 6$ loop positions itself to close in on the active site which also results in the glutamate (Glu141) residue re-positioning for catalytic activity[225] which is clearly visible in Figure 5.15.

5.3.6 Structural comparisons of the archaeal TPIs

All archaeal TPIs are homotetrameric enxymes which can be described as a dimer of dimers. A superposition of the four archaeal TPIs, *Pw*TPI, *Tt*TPI, *Mj*TPI and *Ss*TPI shows that the four structures are very similar, as shown in Figure 5.16.

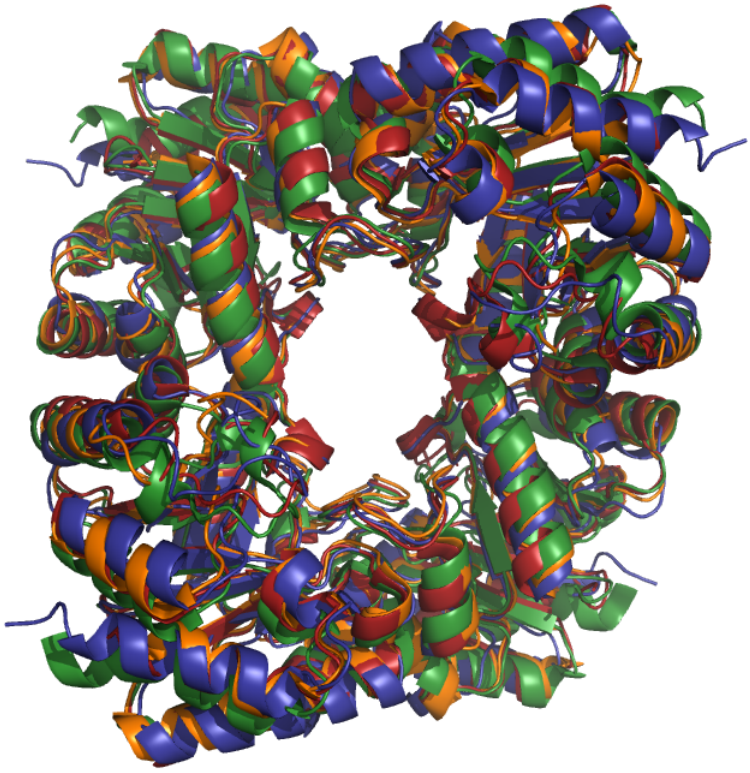


Figure 5.16: Superposition of the archaeal TPI structures with *Pw*TPI (red), *Tt*TPI (green), *Mj*TPI (orange) and *Ss*TPI (blue). Figure was prepared using *PyMOL*.[192]

The rmsd values (calculated using *PyMOL*)[192] are listed below in Table 5.4

rmsd (Å)	*Pw*TPI	*Tt*TPI	*Mj*TPI	*Ss*TPI
*Pw*TPI	-	1.66	0.96	1.85
*Tt*TPI	1.66	-	0.96	2.87
*Mj*TPI	0.96	1.95	-	1.52
*Ss*TPI	1.85	2.87	1.52	-

Table 5.4: Summary of the rmsd values (based on all atoms) between the archaeal TPI tetrameric structures.

The rmsd values show that apart from *Tt*TPI and *Ss*TPI, the structures have rmsd lower than 2 Å. The structures are very similar with no significant changes along the main chain and the larger rmsd values occurring are due to the flexible loops and end terminii. Given the relatively low sequence identities with respect to *Ss*TPI, ranging from 40% to 43%, these results support the notion that protein structure is generally more conserved than the sequence.[230]
*Ss*TPI shows similar features contributing to enzyme stability as observed for other archaeal TPIs. The increased oligomerisation state resulting in the formation of a tetramer is a mechanism employed by archaeal TPIs to increase acid- and thermostability. An increase in the subunit contacts as well as hydrophobic interactions is a result of the higher oligomerisation state of archaeal TPIs. The reduced number of residues compared to other TPIs is also an important factor as is the nature of the residues as it results in tighter packing of the tetramer units (section 5.1.6.2).[210,230]

5.4 Chapter conclusion

Using a standard crystallisation method of setting up trays employing several commercially available crystallisation screens, *Ss*TPI was crystallised and further crystals were grown by using condition optimisation.
Three full datasets were collected and solved for three different structure; two different unbound structures (in the I_{222} and the $P_{2_12_12}$ space groups) and one holo form with PEP bound in the active site (space group $P_{2_12_12}$). PEP is an intermediate in the glycolysis pathway and resembles DHAP, the enzyme substrate.
As expected for a TPI from a hyperthermophilic archaeon, the *Ss*TPI forms tetramers. Comparing the two unbound structures and the PEP bound structure reveals that the $\beta 6\alpha 6$ loop could only be modelled for the PEP bound structure, as the PEP in the active site stabilises it. In crystal form I, the loop appears to be completely open, in crystal form II it appears to be in between the open and closed conforma-

tion, whereas it is completely closed on the active site in the PEP bound structure. Furthermore, the active site Glu141 shifts in position when the ligand is bound in the active site, which is required for catalysis reactions to occur.

Appendices

Appendix A

Additional information for PlcH constructs

A.1 Summary of PlcH primers and vectors

Construct label	Vector	Fusion tag	Parent vector	Forward primer extension	Reverse primer extension
1	pOPINE	...KHHHHHH	pTriEx2	AGGAGATATACC *ATG***ACCGAAA** **ACTGGAAATT** **CCGC**	GTGATGGTGAT GTTTGGTCGCT GCGATGTCGAG
2	pOPINF	MAHHHH HHSSG LEVLFQ GP...	pTriEx2	AAGTTCTGTTT CAGGGCCCGAC CGAAAACTGGA AATTCCGC	ATGGTC*TAG***AAA** **GCTTTAGGTC** **GCTGCGATGT** **CGAG**
3	pOPINE	...KHHHHHH	pTriEx2	AGGAGATATACC *ATG***GAGCCGG** **ACATCCGCACC**	GTGATGGTGAT GTTTGCGTAAC AGCCACACTTCAC
4	pOPINF	MAHHHH HHSSG LEVLFQ GP...	pTriEx2	AAGTTCTGTTT CAGGGCCCGGA GCCGGACATCC GCACC	ATGGTC*TAG***AAA** **GCTTTAGCGT** **AACAGCCACA** **CTTCAC**
5	pOPINE	...KHHHHHH	pTriEx2	AGGAGATATACC *ATG***GAGCCG** **GACATCCGCA** **CC**	GTGATGGTGAT GTTTGCTGGCG GTCACGCTGAG
6	pOPINF	MAHHHH HHSSG LEVLFQ GP...	pTriEx2	AAGTTCTGTTT CAGGGCCCGGA GCCGGACATCC GCACC	ATGGTC*TAG***AAA** **GCTTTAGCTG** **GCGGTCACG** **CTGAG**
7	pOPINE	...KHHHHHH	pTriEx2	AGGAGATATACC *ATG***GAGCCGG** **ACATCCGCACC**	GTGATGGTGAT GTTTGGTCGCT GCGATGTCGAGC
8	pOPINF	MAHHHH HHSSG LEVLFQ GP...	pTriEx2	AAGTTCTGTTT CAGGGCCCGGA GCCGGACATCC GCACC	ATGGTC*TAG***AAA** **GCTTTAGGTC** **GCTGCGATGT** **CGAGC**
9	pOPINE	...KHHHHHH	pTriEx2	AGGAGATATACC *ATG***ATCCAGG** **ACGTCCAGCA** **CG**	GTGATGGTGA TGTTTGTGTCC GGCCAGGCGGC
10	pOPINF	MAHHHH HHSSG LEVLFQ GP...	pTriEx2	AGTTCTGTTTC AGGGCCCGATC CAGGACGTCC AGCACG	ATGGTC*TAG***AAA** **GCTTTAGTGT** **CCGGCCAGG** **CGGC**
11	pOPINE	...KHHHHHH	pTriEx2	AGGAGATATACC *ATG***ATCCAGG** **ACGTCCAGCA** **CG**	GTGATGGTGAT GTTTGACGTTG GTCAGGTCGGG
12	pOPINF	MAHHHH HHSSG LEVLFQ GP...	pTriEx2	AAGTTCTGTTT CAGGGCCCGAT CCAGGACGTCC AGCACG	ATGGTC*TAG***AAA** **GCTTTAGACG** **TTGGTCAGGT** **CGGG**

Table A.1: Summary of the 2×12 primers and the vectors employed for PlcH. Columns summarise construct label, vector, fusion tag, parent vector, DNA sequence of the forward and reverse primers. The primers are written 5' to 3'; the italicised sequence signifies start and stop codons (ATG and TAG), the sequence to the left of the codon is homologous with the plasmid sequence and the sequence in bold is homologous with the construct specific sequence. All vectors carry antibiotic resistance against ampicillin.

Appendix B

Additional data for Chapter 3

B.1 MALDI Mass spectrum of a tryptic digest of PlcH_4

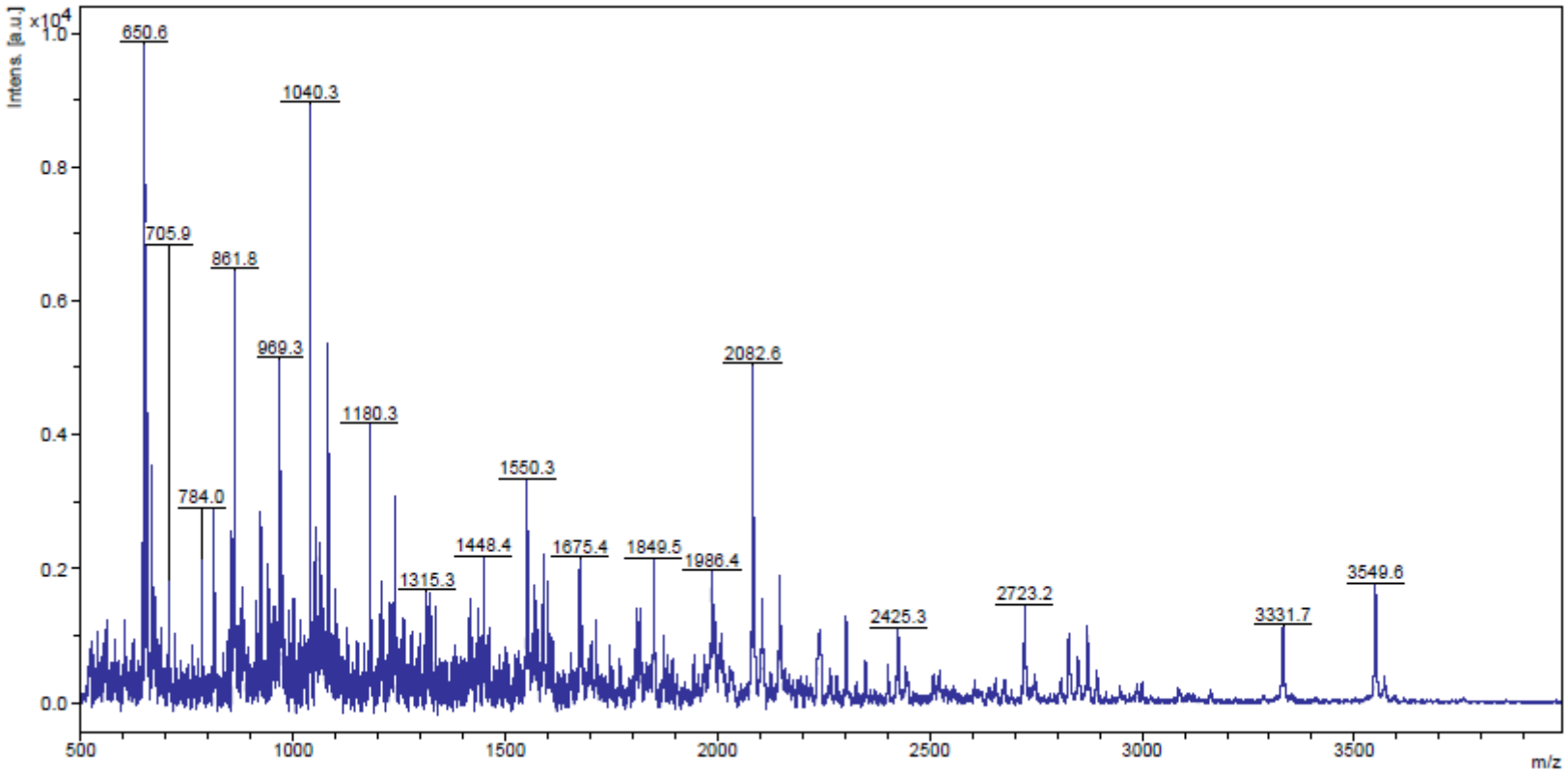


Figure B.1: MALDI Mass spectrum of a tryptic digest of PlcH_4.

B.2 CD data analysis for PlcH_2

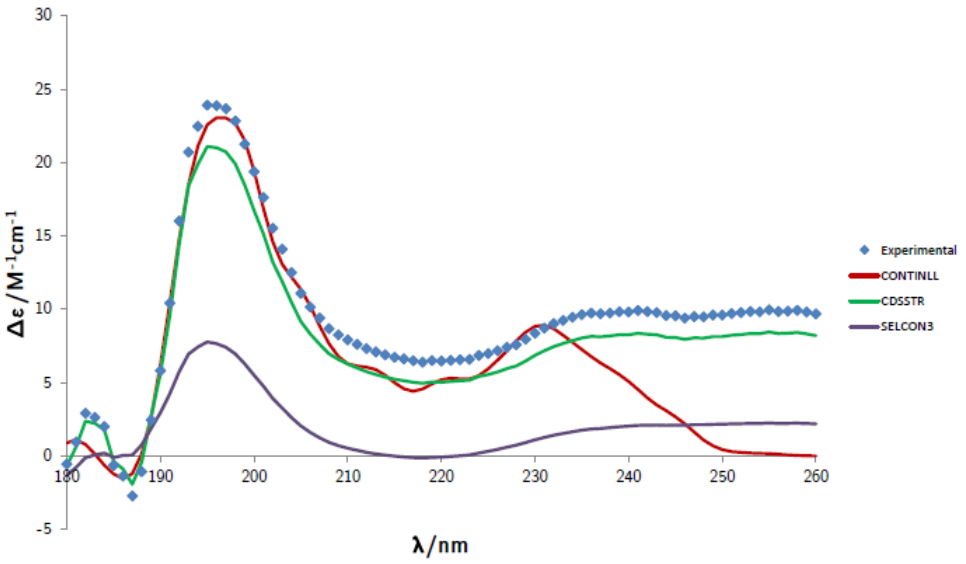


Figure B.2: Experimental CD data and data calculated using the CONTINLL, CDSSTR and SELCON3 programmes, plotted against wavelength, for PlcH_2.

B.3 CD data analysis for PlcH_4

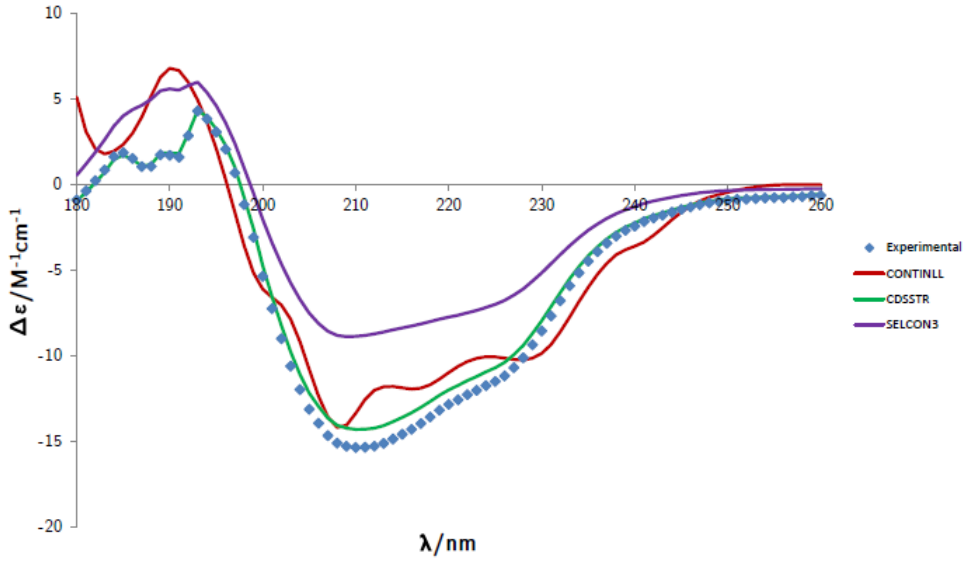


Figure B.3: Experimental CD data and data calculated using the CONTINLL, CDSSTR and SELCON3 programmes, plotted against wavelength, for PlcH_4.

B.4 3D representation of the $PlcHR_2$ activity in the presence of ceramide analogues

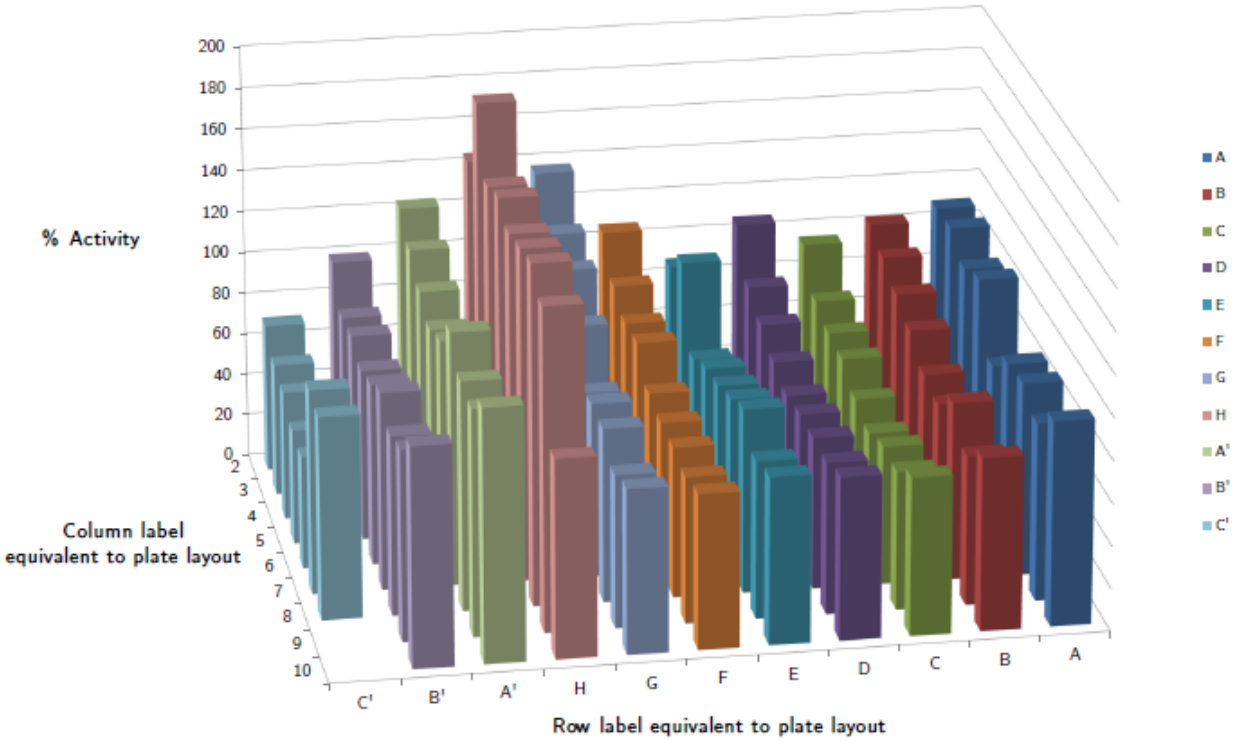


Figure B.4: 3D representation of the percentage activity measured for $PlcHR_2$ in the presence of 10μM of ceramide analogue (at constant protein and substrate concentrations of [$PlcHR_2$] = 0.78 nM and [SM] = 50μM respectively). The layout corresponds to the plate layout shown in B.5. Values are averaged from 3 independent measurements. Error bars omitted for clarity.

B.5 97 ceramide analogues

Ceramide analogues

Plate 1

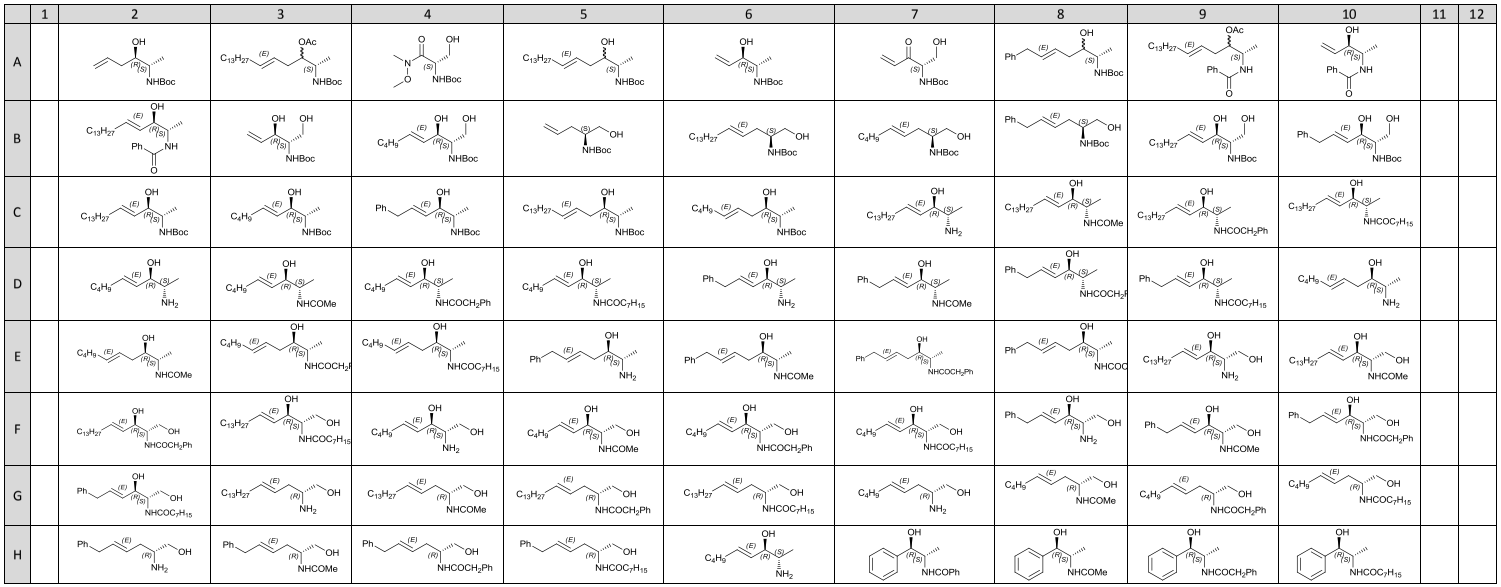
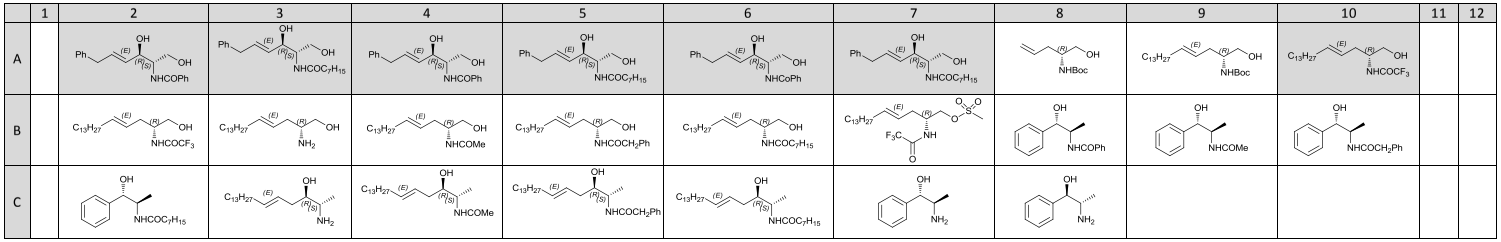


Plate 2



Shaded cells = low purity

Figure B.5: Plate layout showing the structures of all 97 ceramide analogues. The numbers from 1 - 97 correspond to the following key:

Appendix C

Additional data for Chapter 4

C.1 CD data analysis for SPT Δ158

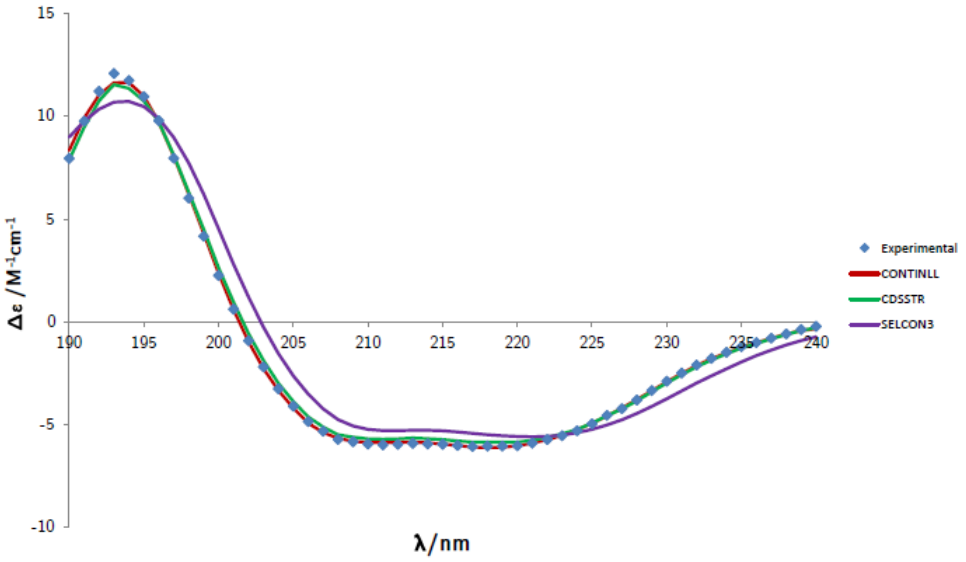


Figure C.1: Experimental CD data and data calculated using the CONTINLL, CDSSTR and SELCON3 programmes, plotted against wavelength, for SPT Δ158.

C.2 CD data analysis for SPT Δ176

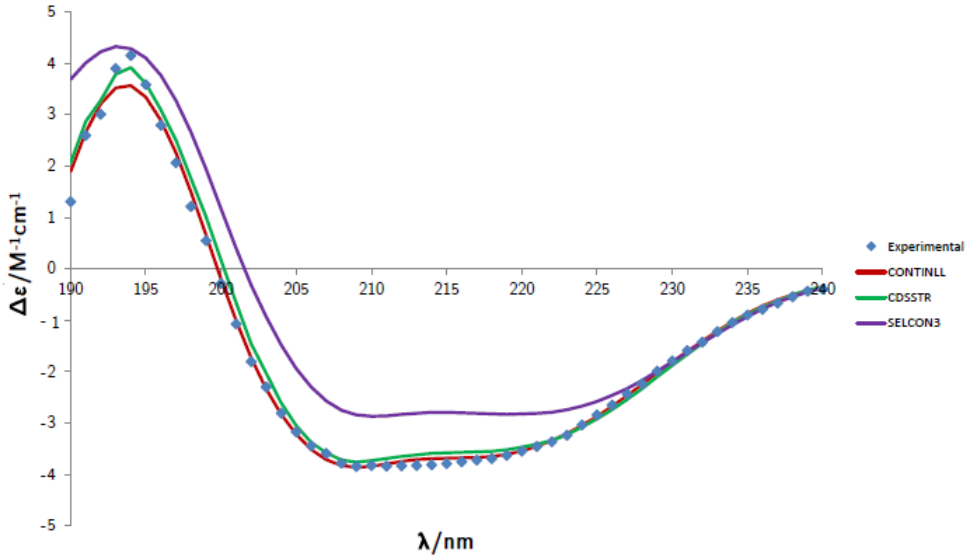


Figure C.2: Experimental CD data and data calculated using the CONTINLL, CDSSTR and SELCON3 programmes, plotted against wavelength, for SPT Δ176.

C.3 CD data analysis for SPT Δ180

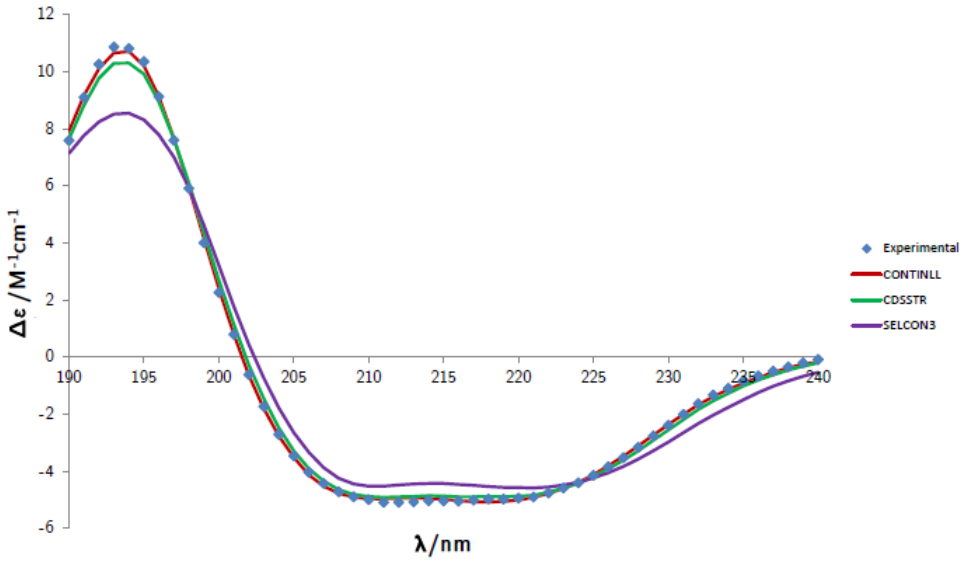


Figure C.3: Experimental CD data and data calculated using the CONTINLL, CDSSTR and SELCON3 programmes, plotted against wavelength, for SPT Δ180.

Bibliography

[1] Gangoiti, P.; Camacho, L.; Arana, L.; Ouro, A.; Granado, M. H.; Brizuela, L.; Casas, J.; Fabriás, G.; Abad, J. L.; Delgado, A.; Gómez-Muñoz, A. *Progress in Lipid Research* **2010**, *49*, 316–334.

[2] Lahiri, S.; Futerman, a. H. *Cellular and molecular life sciences ; CMLS* **2007**, *64*, 2270–2284.

[3] Maréchal, E.; Riou, M.; Kerboeuf, D.; Beugnet, F.; Chaminade, P.; Loiseau, P. M. *Trends in Parasitology* **2011**, *27*, 496–504.

[4] Coppens, I. *Nature Reviews, Microbiology* **2013**, *11*, 823–835.

[5] Carter, H. E.; Norris, W. P.; Glick, J.; Phillips, G. E. *Journal of Biological Chemistry* **1947**, *170*, 269–283.

[6] Young, S. A.; Mina, J. G.; Denny, P. W.; Smith, T. K. *Biochemistry Research International* **2012**, *2012*, Article ID 248135, 12 pages.

[7] Merrill, A. H. *Chemical Reviews* **2011**, *111*, 6387–6422.

[8] Thudichum, J. L. W. *A Treatise on the Chemical Constitution of Brain*; London, 1884.

[9] Pruett, S. T.; Bushnev, A.; Hagedorn, K.; Adiga, M.; Haynes, C. a.; Sullards, M. C.; Liotta, D. C.; Merrill, A. H. *Journal of lipid research* **2008**, *49*, 1621–1639.

[10] Fahy, E.; Subramaniam, S.; Murphy, R. C.; Nishijima, M.; Raetz, C. R. H.; Shimizu, T.; Spener, F.; Meer, G. V.; Wakelam, M. J. O.; Dennis, E. A. *Journal of Lipid Research* **2009**, *50*, S9–S14.

[11] Buccoliero, R.; Futerman, A. H. *Pharmacological Research* **2003**, *47*, 409–419.

[12] Fahy, E.; Cotter, D.; Sud, M.; Subramaniam, S. *Biochimica et Biophysica Acta* **2011**, *1811*, 637–647.

[13] LIPIDMAPS, http://www.lipidmaps.org/.

[14] Fahy, E. et al. *Journal of Lipid Research* **2005**, *46*, 839–861.

[15] Merrill, A. H. *The Journal of Biological Chemistry* **2002**, *277*, 25843–25846.

[16] Karlsson, K. A. *Lipids* **1970**, *5*, 878–891.

[17] Futerman, A. H. *Biochimica et Biophysica Acta* **2006**, *1758*, 1885–1892.

[18] Gault, C. R.; Obeid, L. M.; Hannun, Y. A. *Advances in Experimental Medicinal Biology* **2010**, *688*, 1–23.

[19] Merrill, A. H.; Schmelz, E. M.; Dillehay, D. L.; Spiegel, S.; Shayman, J. a.; Schroeder, J. J.; Riley, R. T.; Voss, K. a.; Wang, E. *Toxicology and Applied Pharmacology* **1997**, *142*, 208–225.

[20] Kolesnick, R. *Journal of Biological Chemistry* **2002**, *110*, 3–8.

[21] Hannun, Y. A.; Obeid, L. M. *The Journal of Biological Chemistry* **2002**, *277*, 25847–25850.

[22] Levade, T.; Malagarie-Cazenave, S.; Gouazé, V.; Ségui, B.; Tardy, C.; Betito, S.; Andrieu-Abadie, N.; Cuvillier, O. *Neurochemical Research* **2002**, *27*, 601–607.

[23] Hannun, Y. A.; Obeid, L. M. *Nature Reviews Molecular Cell Biology* **2008**, *9*, 139–150.

[24] Zhang, K.; Bangs, J. D.; Beverley, S. M. *Advances in Experimental Medicine and Biology* **2010**, *688*, 238–248.

[25] Snell, E. E.; Dimari, S. J.; Brady, R. N. *Chemistry and Physics of Lipids* **1970**, *5*, 116–138.

[26] Stoffel, W. *Chemistry and Physics of Lipids* **1970**, *5*, 139–158.

[27] Hanada, K. *Biochimica et Biophysica Acta* **2003**, *1632*, 16–30.

[28] Luberto, C.; Hannun, Y. A. *The Journal of Biological Chemistry* **1998**, *273*, 14550–14559.

[29] Kolter, T.; Sandhoff, K. *Angewandte Chemie International Edition* **1999**, *38*, 1532–1568.

[30] Vasil, M. L.; Stonehouse, M. J.; Vasil, A. I.; Wadsworth, S. J.; Goldfine, H.; Bolcome, R. E.; Chan, J. *PLoS Pathogens* **2009**, *5*, e1000420, 16 pages.

[31] Bartke, N.; Hannun, Y. A. *Journal of Lipid Research* **2009**, *50*, S91–S96.

[32] Bollinger, C. R.; Teichgräber, V.; Gulbins, E. *Biochimica et Biophysica Acta* **2005**, *1746*, 284–294.

[33] Simons, K.; Toomre, D. *Nature Reviews Molecular Cell Biology* **2000**, *1*, 31–39.

[34] Leslie, M. *Science* **2011**, *334*, 1046–1051.

[35] Toulmay, A.; Prinz, W. A. *The Journal of Cell Biology* **2013**, *202*, 35–44.

[36] Cuvillier, O. *Biochimica et Biophysica Acta* **2002**, *1585*, 153–162.

[37] Albi, E.; Magni, M. V. *The Open Enzyme Inhibition Journal* **2008**, *1*, 72–79.

[38] Berrow, N. S.; Alderton, D.; Sainsbury, S.; Nettleship, J.; Assenberg, R.; Rahman, N.; Stuart, D. I.; Owens, R. J. *Nucleic Acids Research* **2007**, *35*, e45.

[39] Bird, L. E. *Methods* **2011**, *55*, 29–37.

[40] Studier, F. W. *Protein Expression and Purification* **2005**, *41*, 207–234.

[41] Cordingley, M. G.; Register, R. B.; Callahan, P. L.; Garsky, V. M.; Colonno, R. J. *Journal of Virology* **1989**, *63*, 5037–5045.

[42] Sreerama, N.; Woody, R. W. *Analytical Biochemistry* **2000**, *287*, 252–260.

[43] Kelly, S. M.; Jess, T. J.; Price, N. C. *Biochimica et Biophysica Acta* **2005**, *1751*, 119–139.

[44] Greenfield, N. J. *Nature Protocols* **2006**, *1*, 2876–2890.

[45] Whitmore, L.; Wallace, B. A. *Biopolymers* **2008**, *89*, 392–400.

[46] Amplex Red Sphingomyelinase Assay Kit, Manual. 2006.

[47] Zhou, M.; Diwu, Z.; Panchuk-Voloshina, N.; Haugland, R. P. *Analytical Biochemistry* **1997**, *253*, 162–168.

[48] Mohanty, J. G.; Jaffe, J. S.; Schulman, E. S.; Raible, D. G. *Journal of Immunological Methods* **1997**, *202*, 133–141.

[49] Ikushiro, H.; Hayashi, H.; Kagamiyama, H. *The Journal of Biological Chemistry* **2001**, *276*, 18249–18256.

[50] Yard, B. A.; Carter, L. G.; Johnson, K. A.; Overton, I. M.; Dorward, M.; Liu, H.; McMahon, S. A.; Oke, M.; Puech, D.; Barton, G. J.; Naismith, J. H.; Campopiano, D. J. *Journal of Molecular Biology* **2007**, *370*, 870–886.

[51] Niesen, F. H.; Berglund, H.; Vedadi, M. *Nature Protocols* **2007**, *2*, 2212–2221.

[52] Geders, T. W.; Gustafson, K.; Finzel, B. C. *Acta Crystallographica Section F Structural Biology and Crystallization Communications* **2012**, *68*, 596–600.

[53] Pantoliano, M. W.; Petrella, E. C.; Kwasnoski, J. D.; Lobanov, V. S.; Myslik, J.; Graf, E.; Carver, T.; Asel, E.; Springer, B. a.; Lane, P.; Salemme, F. R. *Journal of Biomolecular Screening* **2001**, *6*, 429–440.

[54] Hajizadeh, N. R.; Swann, M. J.; Pohl, E. *Acta Crystallographica Section D Biological Crystallography* **2014**, *Submitted*.

[55] Newman, J.; Egan, D.; Walter, T. S.; Meged, R.; Berry, I.; Ben Jelloul, M.; Sussman, J. L.; Stuart, D. I.; Perrakis, A. *Acta Crystallographica Section D, Biological Crystallography* **2005**, *61*, 1426–1431.

[56] D'Arcy, A.; Mac Sweeney, A.; Stihle, M.; Haber, A. *Acta Crystallographica Section D Biological Crystallography* **2003**, *59*, 396–399.

[57] Gorrec, F. *Journal of Applied Crystallography* **2009**, *42*, 1035–1042.

[58] Brzozowski, A. M.; Walton, J. *Journal of Applied Crystallography* **2001**, *34*, 97–101.

[59] Jancarik, J.; Kim, S. H. *Journal of Applied Crystallography* **1991**, *24*, 409–411.

[60] Wooh, J. W.; Kidd, R. D.; Martin, J. L.; Kobe, B. *Acta Crystallographica Section D Biological Crystallography* **2003**, *D59*, 769–772.

[61] Grimm, C.; Chari, A.; Reuter, K.; Fischer, U. *Acta Crystallographica. Section D, Biological Crystallography* **2010**, *66*, 685–697.

[62] Iglewski, B. H. *Medical Microbiology. 4th edition. Available from: http://www.ncbi.nlm.nih.gov/books/NBK8326/*; 1996; Chapter 27.

[63] Gessard, C. *Comptes rendus de l'Académie des sciences* **1882**, *94*, 536–538.

[64] Stover, C. K. et al. *Nature* **2000**, *406*, 959–964.

[65] Ivanov, I. E.; Kintz, E. N.; Porter, L. a.; Goldberg, J. B.; Burnham, N. A.; Camesano, T. A. *Journal of Bacteriology* **2011**, *193*, 1259–1266.

[66] Ying, Q. L.; Kemme, M.; Simon, S. R. *American Journal of Respiratory Cell and Molecular Biology* **1996**, *15*, 283–291.

[67] Hardalo, C.; Edberg, C. S. *Critical Reviews in Microbiology* **1997**, *23*, 47–75.

[68] Yamanaka, T.; Kijimoto, S.; Okunuki, K. *The Journal of Biochemistry* **1963**, *33*, 416–421.

[69] Masuda, N.; Sakagawa, E.; Ohya, S. *Antimicrobial Agents and Chemotherapy* **1995**, *39*, 645–649.

[70] Hancock, R. E. *Clinical Infectious Diseases* **1998**, *27 (Suppl*, S93–S99.

[71] Filloux, A.; Bally, M.; Ball, G.; Akrim, M.; Tommassen, J.; Lazdunski, A. *The EMBO Journal* **1990**, *9*, 4323–4329.

[72] Filloux, A. *Frontiers in Microbiology* **2011**, *2*, Article 155, 21 pages.

[73] Sargent, F.; Ballantine, S. P.; Rugman, P. A.; Palmer, T.; Boxer, D. H. *European journal of biochemistry, FEBS* **1998**, *255*, 746–754.

[74] Berks, B. C.; Sargent, F.; Palmer, T. *Molecular Microbiology* **2000**, *35*, 260–274.

[75] Snyder, A.; Vasil, A. I.; Zajdowicz, S. L.; Wilson, Z. R.; Vasil, M. L. *Journal of Bacteriology* **2006**, *188*, 1762–1774.

[76] Aiello, D.; Williams, J. D.; Majgier-Baranowska, H.; Patel, I.; Peet, N. P.; Huang, J.; Lory, S.; Bowlin, T. L.; Moir, D. T. *Antimicrobial Agents and Chemotherapy* **2010**, *54*, 1988–1999.

[77] Galle, M.; Carpentier, I.; Beyaert, R. *Current Protein and Peptide Science* **2012**, *13*, 831–842.

[78] Vento, S.; Cainelli, F.; Temesgen, Z. *The Lancet Oncology* **2008**, *9*, 982–992.

[79] Rada, B.; Gardina, P.; Myers, T. G.; Leto, T. L. *Mucosal Immunology* **2011**, *4*, 158–171.

[80] Goldberg, J. B. *F1000 Biology Reports* **2010**, *2*, 27–30.

[81] El Solh, A. A.; Akinnusi, M. E.; Wiener-Kronish, J. P.; Lynch, S. V.; Pineda, L. A.; Szarpa, K. *American Journal of Respiratory and Critical Medical Care* **2008**, *178*, 513–519.

[82] Hirche, T. O.; Benabid, R.; Deslee, G.; Gangloff, S.; Achilefu, S.; Guenounou, M.; Lebargy, F.; Hancock, R. E.; Belaaouaj, A. *The Journal of Immunology* **2008**, *181*, 4945–4954.

[83] Hø iby, N. *BMC Medicine* **2011**, *9*, 32 (7 pages).

[84] Ramu, Y.; Xu, Y.; Lu, Z. *Proceedings of the National Academy of Sciences; PNAS* **2007**, *104*, 6448–6453.

[85] Hansen, C. R.; Pressler, T.; Hø iby, N. *Journal of Cystic Fibrosis* **2008**, *7*, 523–530.

[86] Stonehouse, M. J.; Cota-Gomez, A.; Parker, S. K.; Martin, W. E.; Hankin, J. A.; Murphy, R. C.; Chen, W.; Lim, K. B.; Hackett, M.; Vasil, A. I.; Vasil, M. L. *Molecular Microbiology* **2002**, *46*, 661–676.

[87] Koch, C. *Pediatric Pulmonology* **2002**, *34*, 232–236.

[88] Liffourrena, A. S.; Massimelli, M. J.; Forrellad, M. A.; Lisa, A. T.; Domenech, C. E.; Lucchesi, G. I. *Current Microbiology* **2007**, *55*, 530–536.

[89] Gaman, W.; Cates, C.; Snelling, C. F. T.; Lank, B.; Ronald, A. R. *Antimicrobial Agents and Chemotherapy* **1976**, *9*, 474–480.

[90] Touhami, A.; Jericho, M. H.; Boyd, J. M.; Beveridge, T. J. *Journal of Bacteriology* **2006**, *188*, 370–377.

[91] Ramphal, R.; Pier, G. B. *Infection and Immunity* **1985**, *47*, 1–4.

[92] Hø iby, N.; Ciofu, O.; Johansen, H. K.; Song, Z.-j.; Moser, C.; Jensen, P. O.; Molin, S.; Givskov, M.; Tolker-Nielsen, T.; Bjarnsholt, T. *International Journal of Oral Science* **2011**, *3*, 55–65.

[93] Shaver, C. M.; Hauser, A. R. *Infection and Immunity* **2004**, *72*, 6969–6977.

[94] Callahan III, L. T. *Infection and Immunity* **1976**, *14*, 55–61.

[95] Rocha, C. L.; Coburn, J.; Rucks, E. A.; Olson, J. C. *Infection and Immunity* **2003**, *71*, 5296–5305.

[96] Pritchard, A. E.; Vasil, M. L. *Journal of Bacteriology* **1986**, *167*, 291–298.

[97] Cotar, A.-I.; Chifiriuc, M.-C.; Dinu, S.; Bucur, M.; Iordache, C.; Banu, O.; Dracea, O.; Larion, C.; Lazar, V. *International Journal of Molecular Sciences* **2010**, *11*, 5273–5291.

[98] Titball, R. W. *Microbiology Reviews* **1993**, *57*, 347–366.

[99] Vasil, M. L. *Pseudomonas* **2006**, *4*, 69–98.

[100] Sato, H.; Frank, D. W.; Hillard, C. J.; Feix, J. B.; Pankhaniya, R. R.; Moriyama, K.; Finck-barbanc, V.; Buchaklian, A.; Lei, M.; Long, R. M.; Wiener-kronish, J.; Sawa, T. *The EMBO Journal* **2003**, *22*, 2959–2969.

[101] Wilderman, P. J.; Vasil, A. I.; Johnson, Z.; Vasil, M. L. *Molecular Microbiology* **2001**, *39*, 291–303.

[102] Hinnebusch, B. J.; Rudolph, A. E.; Cherepanov, P.; Dixon, J. E.; Schwan, T. G.; Forsberg, A. *Science* **2002**, *296*, 733–735.

[103] Macfarlane, M. G.; Knight, B. C. *Biochemical Journal* **1941**, *35*, 884–902.

[104] Johansen, K. A.; Gill, R. E.; Vasil, M. L.; Johansen, K. A.; Gill, R. E. *Infection and Immunity* **1996**, *64*, 3259–3266.

[105] Liu, P. V. *Journal of Infectious Diseases* **1966**, *116*, 112–116.

[106] Vasil, M. L.; Berka, R. M.; Gray, G. L.; Nakai, H. *Journal of Bacteriology* **1982**, *152*, 431–440.

[107] Luberto, C.; Stonehouse, M. J.; Collins, E. A.; Marchesini, N.; El-Bawab, S.; Vasil, A. I.; Vasil, M. L.; Hannun, Y. A. *The Journal of Biological Chemistry* **2003**, *278*, 32733–32743.

[108] Terada, L. S.; Johansen, K. A.; Nowbar, S.; Vasil, A. I.; Vasil, M. L. *Infection and Immunity* **1999**, *67*, 2371–2376.

[109] Ostroff, R. M.; Vasil, M. L. *Journal of Bacteriology* **1987**, *169*, 4597–4601.

[110] Ostroff, R. M.; Vasil, A. I.; Vasil, M. L. *Journal of Bacteriology* **1990**, *172*, 5915–5923.

[111] Cota-Gomez, A.; Vasil, A. I.; Kadurugamuwa, J.; Beveridge, T. J.; Schweizer, H. P.; Vasil, M. L. *Infection and Immunity* **1997**, *65*, 2904–2913.

[112] Montes, L.-R.; López, D. J.; Sot, J.; Bagatolli, L. A.; Stonehouse, M. J.; Vasil, M. L.; Wu, B. X.; Hannun, Y. A.; Goñi, F. M.; Alonso, A. *Biochemistry* **2008**, *47*, 11222–11230.

[113] Zhang, Y. *BMC Bioinformatics* **2008**, *9*, 40 (8 pages).

[114] Kadurugamuwa, J. L.; Beveridge, T. J. *Journal of Bacteriology* **1995**, *177*, 3398–4008.

[115] Ibarguren, M.; Bomans, P. H. H.; Frederik, P. M.; Stonehouse, M.; Vasil, A. I.; Vasil, M. L.; Alonso, A.; Goñi, F. M. *Biochimica et Biophysica Acta* **2010**, *1798*, 59–64.

[116] Exton, J. H. *Biochimica et Biophysica Acta* **1994**, *1212*, 26–42.

[117] Truan, D.; Vasil, A.; Stonehouse, M.; Vasil, M. L.; Pohl, E. *Protein Expression and Purification* **2013**, *90*, 40–46.

[118] Felts, R. L.; Reilly, T. J.; Tanner, J. J. *The Journal of Biological Chemistry* **2006**, *281*, 30289–30298.

[119] Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G. *Molecular Systems Biology* **2011**, *7*, Article 539, 7 pages.

[120] Goujon, M.; McWilliam, H.; Li, W.; Valentin, F.; Squizzato, S.; Paern, J.; Lopez, R. *Nucleic Acids Research* **2010**, *38*, W695–W699.

[121] Gouet, P. *Nucleic Acids Research* **2003**, *31*, 3320–3323.

[122] Vedadi, M. et al. *Proceedings of the National Academy of Sciences; PNAS* **2006**, *103*, 15835–15840.

[123] Wargo, M. J.; Gross, M. J.; Rajamani, S.; Allard, J. L.; Lundblad, L. K. A.; Allen, G. B.; Vasil, M. L.; Leclair, L. W.; Hogan, D. A. *American Journal of Respiratory and Critical Care Medicine* **2011**, *184*, 345–354.

[124] Battye, T. G. G.; Kontogiannis, L.; Johnson, O.; Powell, H. R.; Leslie, A. G. W. *Acta Crystallographica. Section D, Biological Crystallography* **2011**, *67*, 271–281.

[125] Mohanty, B. K.; Maples, V. F.; Kushner, S. R. *Molecular Microbiology* **2004**, *54*, 905–920.

[126] McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *Journal of Applied Crystallography* **2007**, *40*, 658–674.

[127] Mina, J. G.; Mosely, J. A.; Ali, H. Z.; Denny, P. W.; Steel, P. G. *Organic & Biomolecular Chemistry* **2011**, *9*, 1823–1830.

[128] Yedgar, S.; Barenholz, Y.; Copper, V. G. *Biochimica et Biophysica Acta* **1974**, *363*, 98–111.

[129] *European Journal of Biochemistry* **1979**, *97*, 319–320.

[130] Miroux, B.; Walker, J. E. *Journal of Molecular Biology* **1996**, *260*, 289–298.

[131] Dumon-Seignovert, L.; Cariot, G.; Vuillard, L. *Protein Expression and Purification* **2004**, *37*, 203–206.

[132] ArcticExpress Competent Cells and ArcticExpress (DE3) Competent Cells Instruction Manual. 2010.

[133] Bolanos-Garcia, V. M.; Davies, O. R. *Biochimica et Biophysica Acta* **2006**, *1760*, 1304–1313.

[134] Young, S. A.; Smith, T. K. *Molecular Microbiology* **2010**, *76*, 1461–1482.

[135] Carman, G. M.; Deems, R. A.; Dennis, E. A. *The Journal of Biological Chemistry* **1995**, *270*, 18711–18714.

[136] Deems, R. A.; Eaton, B. R.; Dennis, E. A. *The Journal of Biological Chemistry* **1975**, *250*, 9013–9020.

[137] Yedgar, S.; Gatt, S. *Biochemistry* **1976**, *15*, 2570–2573.

[138] London, E.; Brown, D. A. *Biochimica et Biophysica Acta* **2000**, *1508*, 182–195.

[139] Deems, R. A. *Analytical Biochemistry* **2000**, *287*, 1–16.

[140] Hill, D.; Dubey, J. P. *Clinical Microbiology and Infection* **2002**, *8*, 634–640.

[141] Dubey, J. P.; Weigel, R. M.; Siegel, A. M.; Thulliez, P.; Kitron, U. D.; Mitchell, M. A.; Mannelli, A.; Mateus-Pinilla, N. E.; Shen, S. K.; Kwok, O. C. *The Journal of Parasitology* **1995**, *81*, 723–729.

[142] Weissmann, J. *The Canadian Veterinary Journal* **2003**, *44*, 322–324.

[143] Kim, K.; Weiss, L. M. *International Journal for Parasitology* **2004**, *34*, 423–432.

[144] Factsheet on the World Malaria Report 2013; http://www.who.int/malaria/media/world_malaria_report_2013/en/.

[145] Black, M. W.; Boothroyd, J. C. *Microbiology and Molecular Biology Reviews ; MMBR* **2000**, *64*, 607–623.

[146] Sweeney, K. R.; Morrissette, N. S.; LaChapelle, S.; Blader, I. J. *Eukaryotic Cell* **2010**, *9*, 1680–1689.

[147] Innes, E. A.; Bartley, P. M.; Maley, S.; Katzer, F.; Buxton, D. *Memórias do Instituto Oswaldo Cruz* **2009**, *104*, 246–251.

[148] Dubey, J. P. *The Journal of Eukaryotic Microbiology* **2008**, *55*, 467–475.

[149] Dubey, J. *International Journal for Parasitology* **2009**, *39*, 877–882.

[150] Tenter, A. M.; Heckeroth, A. R.; Weiss, L. M. *International Journal for Parasitology* **2000**, *30*, 1217–1258.

[151] Frenkel, J. K.; Dubey, J. P.; Miller, N. L. *Science* **1970**, *167*, 893–896.

[152] Dubey, J. P.; Lindsay, D. S.; Speer, C. A. *Clinical Microbiology Reviews* **1998**, *11*, 267–299.

[153] Kappe, S. H.; Buscaglia, C. A.; Bergman, L. W.; Coppens, I.; Nussenzweig, V. *Trends in Parasitology* **2004**, *20*, 13–16.

[154] Hå kansson, S.; Charron, A. J.; Sibley, L. D. *The EMBO Journal* **2001**, *20*, 3132–3144.

[155] Di Cristina, M.; Marocco, D.; Galizi, R.; Proietti, C.; Spaccapelo, R.; Crisanti, A. *Infection and Immunity* **2008**, *76*, 3491–3501.

[156] Gross, U.; Bohne, W.; Soête, M.; Dubremetz, J. F. *Parasitology Today* **1996**, *12*, 30–33.

[157] Skariah, S.; Mcintyre, M. K.; Mordue, D. G. *Parasitology Research* **2010**, *107*, 253–260.

[158] Feustel, S. M.; Meissner, M.; Liesenfeld, O. *Virulence* **2012**, *3*, 182–192.

[159] Lachenmaier, S. M.; Deli, M. A.; Meissner, M.; Liesenfeld, O. *Journal of Neuroimmunology* **2011**, *232*, 119–130.

[160] Jones, J. L.; Dubey, J. P. *Clinical Infectious Diseases, Food Safety* **2012**, *55*, 845–851.

[161] Cook, A. J. C.; Gilbert, R. E.; Buffolano, W.; Zufferey, J.; Petersen, E.; Jenum, P. A.; Foulon, W.; Semprini, A. E. *BMJ* **2000**, *321*, 142–147.

[162] Hanauer, D. A.; Ramakrishnan, N.; Seyfried, L. S. *PLoS ONE* **2013**, *8*, e70585, 10 pages.

[163] Webster, J. P. *Schizophrenia Bulletin* **2007**, *33*, 752–756.

[164] Worth, A. R.; Lymbery, A. J.; Thompson, R. C. A. *Trends in Parasitology* **2013**, *29*, 150–155.

[165] Alvarado-Esquivel, C.; Campillo-Ruiz, F.; Liesenfeld, O. *Parasites & Vectors* **2013**, *6*, 113, 6 pages.

[166] McFadden, D. C.; Camps, M.; Boothroyd, J. C. *Drug Resistance Updates* **2001**, *4*, 79–84.

[167] Sonda, S.; Hehl, A. B. *Trends in Parasitology* **2006**, *22*, 41–47.

[168] Azzouz, N.; Rauscher, B.; Gerold, P.; Cesbron-Delauw, M.-F.; Dubremetz, J.-F.; Schwarz, R. T. *International Journal for Parasitology* **2002**, *32*, 677–684.

[169] de Melo, E. J.; de Souza, W. *Cell Structure and Function* **1996**, *21*, 47–52.

[170] Bisanz, C.; Bastien, O.; Grando, D.; Jouhet, J.; Maréchal, E.; Cesbron-Delauw, M.-F. *The Biochemical Journal* **2006**, *394*, 197–205.

[171] Coppens, I. *Cellular Microbiology* **2006**, *8*, 1–9.

[172] Sonda, S.; Sala, G.; Ghidoni, R. *Antimicrobial Agents* **2005**, *49*, 1794–1801.

[173] Pratt, S.; Wansadhipathi-Kannangara, N. K.; Bruce, C. R.; Mina, J. G.; Shams-Eldin, H.; Casas, J.; Hanada, K.; Schwarz, R. T.; Sonda, S.; Denny, P. W. *Molecular and Biochemical Parasitology* **2013**, *187*, 43–51.

[174] Hanada, K.; Hara, T.; Nishijima, M. *The Journal of Biological Chemistry* **2000**, *275*, 8409–8415.

[175] Yasuda, S.; Nishijima, M.; Hanada, K. *The Journal of Biological Chemistry* **2003**, *278*, 4176–4183.

[176] Ikushiro, H.; Islam, M. M.; Okamoto, A.; Hoseki, J.; Murakawa, T.; Fujii, S.; Miyahara, I.; Hayashi, H. *The Journal of Biochemistry* **2009**, *146*, 549–562.

[177] Hornemann, T.; Wei, Y.; von Eckardstein, A. *The Biochemical Journal* **2007**, *405*, 157–164.

[178] Ikushiro, H.; Islam, M. M.; Tojo, H.; Hayashi, H. *Journal of Bacteriology* **2007**, *189*, 5749–5761.

[179] Shiraiwa, Y.; Ikushiro, H.; Hayashi, H. *The Journal of Biological Chemistry* **2009**, *284*, 15487–15495.

[180] Ikushiro, H.; Fujii, S.; Shiraiwa, Y.; Hayashi, H. *The Journal of Biological Chemistry* **2008**, *283*, 7542–7553.

[181] Hanada, K.; Nishijima, M.; Fujita, T.; Kobayashi, S. *Biochemical Pharmacology* **2000**, *59*, 1211–1216.

[182] Lowther, J.; Yard, B. A.; Johnson, K. A.; Carter, L. G.; Bhat, V. T.; Raman, M. C. C.; Clarke, D. J.; Ramakers, B.; McMahon, S. A.; Naismith, J. H.; Campopiano, D. J. *Molecular Biosystems* **2010**, *6*, 1682–1693.

[183] Miyake, Y.; Kozutsumi, Y.; Nakamura, S.; Fujita, T.; Kawasaki, T. *Biochemical and Biophysical Research Communications* **1995**, *211*, 396–403.

[184] Wadsworth, J. M.; Clarke, D. J.; McMahon, S. A.; Lowther, J. P.; Beattie, A. E.; Langridge-Smith, P. R. R.; Broughton, H. B.; Dunn, T. M.; Naismith, J. H.; Campopiano, D. J. *Journal of the American Chemical Society* **2013**, *135*, 14276–14285.

[185] Ikushiro, H.; Hayashi, H.; Kagamiyama, H. *Biochemistry* **2004**, *43*, 1082–1092.

[186] Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *Journal of Molecular Biology* **1990**, *215*, 403–410.

[187] Hofmann, K.; Stoffel, W. *Biological Chemistry Hoppe.Seyler* **1993**, *374*, 166–171.

[188] Cole, C.; Barber, J. D.; Barton, G. J. *Nucleic Acids Research* **2008**, *36*, W197–W201.

[189] Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G. *Bioinformatics* **2007**, *23*, 2947–2948.

[190] Šali, A.; Blundell, T. L. *Journal of Molecular Biology* **1993**, *234*, 779–815.

[191] Emsley, P.; Cowtan, K. *Acta Crystallographica. Section D, Biological Crystallography* **2004**, *60*, 2126–2132.

[192] DeLano, W. L. The PyMOL Molecular Graphics System; http://www.pymol.org. 2002.

[193] Krissinel, E.; Henrick, K. *Journal of Molecular Biology* **2007**, *372*, 774–797.

[194] Blanchet, C. E.; Svergun, D. I. *Annual Review of Physical Chemistry* **2013**, *64*, 37–54.

[195] McPherson, A. *Protein Science* **2001**, 418–422.

[196] McInerney, J. O. *Bioinformatics* **1998**, *14*, 372–373.

[197] Rütti, M. F.; Richard, S.; Penno, A.; von Eckardstein, A.; Hornemann, T. *Journal of Lipid Research* **2009**, *50*, 1237–1244.

[198] Riley, R. T.; Norred, W. P.; Wang, E.; Merrill, A. H. *Natural Toxins* **1999**, *7*, 407–414.

[199] Zillig, W.; Stetter, K. O.; Wunderl, S.; Schulz, W.; Priess, H.; Scholz, I. *Archives of Microbiology* **1980**, *125*, 259–269.

[200] Brock, T. D.; Freeze, H. *Journal of Bacteriology* **1969**, *98*, 289–297.

[201] Madigan, M. T.; Marrs, B. L. *Scientific American* **1997**, 83–87.

[202] Brouns, S. J. J.; Ettema, T. J. G.; Stedman, K. M.; Walther, J.; Smidt, H.; Snijders, A. P. L.; Young, M.; Bernander, R.; Wright, P. C.; Siebers, B.; van der Oost, J. *Geothermal Biology and Geochemistry in Yellowstone National Park* **2005**, 261–276.

[203] Brock, T. D.; Brock, K. M.; Belly, R. T.; Weiss, R. L. *Archives of Microbiology* **1972**, *84*, 54–68.

[204] Woese, C. R. *Microbiological Reviews* **1987**, *51*, 221–271.

[205] Pace, N. R. *Science* **1997**, *276*, 734–740.

[206] Daniel, R. M.; Cowan, D. A. *Cellular and Molecular Life Sciences; CMLS* **2000**, *57*, 250–264.

[207] Grogan, D. W. *Journal of Bacteriology* **1989**, *171*, 6710–6719.

[208] Zillig, W.; Arnold, H. P.; Holz, I.; Prangishvili, D.; Schweier, A.; Stedman, K.; She, Q.; Phan, H.; Garrett, R.; Kristjansson, J. K. *Extremophiles* **1998**, *2*, 131–140.

[209] She, Q. et al. *Proceedings of the National Academy of Sciences; PNAS* **2001**, *98*, 7835–7840.

[210] Walden, H.; Bell, G. S.; Russell, R. J.; Siebers, B.; Hensel, R.; Taylor, G. L. *Journal of Molecular Biology* **2001**, *306*, 745–757.

[211] Banner, D. W.; Bloomer, A. C.; Petsko, G. A.; Phillips, D. C.; Pogson, C. I.; Wilson, I. A.; Corran, P. H.; Furth, A. J.; Milman, J. D.; Offord, R. E.; Priddle, J. D.; Waley, S. G. *Nature* **1975**, *255*, 609–614.

[212] Sterner, R.; Höcker, B. *Chemical Reviews* **2005**, *105*, 4038–4055.

[213] Wierenga, R. K. *FEBS Letters* **2001**, *492*, 193–198.

[214] Höcker, B. *Biomolecular Engineering* **2005**, *22*, 31–38.

[215] Vega, M.; Lorentzen, E.; Linden, A.; Willmans, M. *Current Opinion in Chemical Biology* **2003**, *7*, 694–701.

[216] Nagano, N.; Orengo, C. A.; Thornton, J. M. *Journal of Molecular Biology* **2002**, *321*, 741–765.

[217] Raychaudhuri, S.; Younas, F.; Karplus, P. A.; Faerman, C. H.; Ripoll, D. R. *Protein Science* **1997**, *6*, 1849–1857.

[218] Meyerhof, O.; Kiessling, W. *Biochemische Zeitschrift* **1935**, *279*, 40–48.

[219] Walden, H.; Taylor, G. L.; Lorentzen, E.; Pohl, E.; Lilie, H.; Schramm, A.; Knura, T.; Stubbe, K.; Tjaden, B.; Hensel, R. *Journal of Molecular Biology* **2004**, *342*, 861–875.

[220] Guallar, V.; Jacobson, M.; McDermott, A.; Friesner, R. A. *Journal of Molecular Biology* **2004**, *337*, 227–239.

[221] Knowles, J. R. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **1991**, *332*, 115–121.

[222] Harris, T. K.; Abeygunawardana, C.; Mildvan, A. S. *Biochemistry* **1997**, *36*, 14661–14675.

[223] Kursula, I.; Wierenga, R. K. *The Journal of Biological Chemistry* **2003**, *278*, 9544–9551.

[224] Harris, T. K.; Cole, R. N.; Comer, F. I.; Mildvan, A. S. *Biochemistry* **1998**, *37*, 16828–16838.

[225] Xiang, J.; Jung, J.-Y.; Sampson, N. S. *Biochemistry* **2004**, *43*, 11436–11445.

[226] Esser, D.; Kouril, T.; Talfournier, F.; Polkowska, J.; Schrader, T.; Bräsen, C.; Siebers, B. *Extremophiles* **2013**, *17*, 205–216.

[227] Zaparty, M. et al. *Extremophiles* **2010**, *14*, 119–142.

[228] Lambeir, A.-M.; Opperdoes, F. R.; Pathology, M.; Diseases, T.; Physics, C.; February, R. *European Journal of Biochemistry, FEBS* **1987**, *168*, 69–74.

[229] Tomlinson, I. D.; Turner, J. F. *Phytochemistry* **1979**, *18*, 1959–1962.

[230] Lorentzen, E.; Pohl, E.; Zwart, P.; Stark, A.; Russell, R. B.; Knura, T.; Hensel, R.; Siebers, B. *The Journal of Biological Chemistry* **2003**, *278*, 47253–47260.

[231] Copley, R. R.; Bork, P. *Journal of Molecular Biology* **2000**, *303*, 627–641.

[232] Farber, G. K.; Petsko, G. A. *Trends in Biochemical Sciences* **1990**, *15*, 228–234.

[233] Jaenicke, R.; Böhm, G. *Current Oinion in Structural Biology* **1998**, *8*, 738–748.

[234] Kumar, S.; Nussinov, R. *Cellular and molecular life sciences; CMLS* **2001**, *58*, 1216–1233.

[235] Andreotti, G.; Tutino, M. L.; Sannia, G.; Marino, G.; Cubellis, M. V. *Biochimica et Biophysica Acta* **1994**, *1208*, 310–315.

[236] Andreotti, G.; Cubellis, M. V.; Di Palo, M.; Fessas, D.; Sannia, G.; Marino, G. *Biochemical Journal* **1997**, *264*, 259–264.

[237] Grutter, M. G.; Hawkes, R. B.; Matthews, B. W. *Nature* **1979**, *277*, 667–669.

[238] Daniel, R. M.; Dines, M.; Petach, H. H. *The Biochemical Journal* **1996**, *317*, 1–11.

[239] Shoichet, K.; Baase, W. A.; Kuroki, R.; Matthews, B. W. *Proceedings of the National Academy of Sciences; PNAS* **1995**, *92*, 452–456.

[240] Gayathri, P.; Banerjee, M.; Vijayalakshmi, A.; Azeez, S.; Balaram, H.; Balaram, P.; Murthy, M. R. N. *Acta Crystallographica. Section D, Biological Crystallography* **2007**, *63*, 206–220.

[241] Xiao, L.; Honig, B. *Journal of Molecular Biology* **1999**, *289*, 1435–1444.

[242] Dams, T.; Auerbach, G.; Bader, G.; Jacob, U.; Ploom, T.; Huber, R.; Jaenicke, R. *Journal of Molecular Biology* **2000**, *297*, 659–672.

[243] Kiefer, F.; Arnold, K.; Künzli, M.; Bordoli, L.; Schwede, T. *Nucleic Acids Research* **2009**, *37*, D387–D392.

[244] Knobeloch, D.; Schmidt, D.; Scheerer, P.; Krauss, N.; Wessner, H.; Scholz, C.; Küttner,; Von Rintelen, T.; Wessel, A.; Höhne, W. *Insect Molecular Biology* **2010**, *19*, 35–48.

[245] Broennimann, C.; Eikenberry, E. F.; Henrich, B.; Horisberger, R.; Huelsen, G.; Pohl, E.; Schmitt, B.; Schulze-Briese, C.; Suzuki, M.; Tomizaki, T.; Toyokawa, H.; Wagner, A. *Journal of Synchrotron Radiation* **2006**, *13*, 120–130.

[246] Winn, M. D. et al. *Acta Crystallographica. Section D, Biological Crystallography* **2011**, *67*, 235–242.

[247] Adams, P. D. et al. *Acta crystallographica. Section D, Biological Crystallography* **2010**, *66*, 213–221.

[248] Lovell, S. C.; Davis, I. W.; Arendall III, W. B.; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. *Proteins* **2003**, *450*, 437–450.

[249] Wlodawer, A.; Minor, W.; Dauter, Z.; Jaskolski, M. *The FEBS Journal* **2008**, *275*, 1–21.

[250] Grüning, N.-M.; Du, D.; Keller, M. A.; Luisi, B. F.; Ralser, M. *Open Biology* **2014**, *4*, 130232, 12 pages.

[251] Joseph, D.; Petsko, G. A.; Karplus, M. *Science* **1990**, *249*, 1425–1428.